

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



JUMPING GENOMES

Genomic evolution in the palaeotetraploid lab-model frog *Xenopus laevis* **PAGES 320 & 336**

SOCIAL SCIENCE

WHAT'S UP WITH THE POLLS?

Why the Trump/Clinton election is so tough to predict

PAGE 304

ARTIFICIAL INTELLIGENCE

MISSING THE POINT ON AI

Autonomous systems have social costs now

PAGE 311

ASTROPHYSICS

A FLARE FOR MYSTERY

Unexplained X-ray sources found in archival data

PAGES 321 & 356

NATURE.COM/NATURE

20 October 2016 £10

Vol. 538, No. 7625



THIS WEEK

EDITORIALS

EQUALITY A tool to help fight for more female speakers at conferences **p.290**

WORLD VIEW Consciousness is a red herring in AI concerns **p.291**



PAIN FREE How the naked mole rat ignores burning heat **p.293**

Let reason prevail

After an election campaign like no other, Hillary Clinton will make a fine US president, and not only because she is not Donald Trump.

In March 2011, this publication suggested that the US Congress seemed lost in the “intellectual wilderness”. The Republicans had taken over the House of Representatives, and one of the early acts of the chamber’s science committee was to approve legislation that denied the threat of climate change. As it turns out, this was just one tiny piece of a broader populist movement that was poised to transform the US political scene. Judging by the current presidential campaign, when it comes to reason, decency and use of evidence, much of the country’s political system seems to have lost its way.

Is there anything left to say about the unsuitability of Donald Trump as a presidential candidate? Even senior figures of his own party have disowned him. The latest revelations about his sordid attitude and behaviour towards women only confirm what was obvious to many from the very beginning: Trump is a demagogue not fit for high office, or for the responsibilities that come with it.

Will the centre hold? Will the United States elect its first female president, Hillary Clinton? It should do. And not just because she is not Donald Trump. Clinton is a quintessential politician — and a good one at that. She has shown tremendous understanding of complex issues directly relevant to *Nature*’s readers, and has engaged with scientists and academics. Take health: as first lady, she led attempts to expand health care in the early years of her husband Bill Clinton’s presidency. She supported the Children’s Health Insurance Program, which reaches millions of poor children. She championed women’s rights, and as secretary of state made global health a priority through the Global Health Initiative, a framework to coordinate various US programmes. Clinton may not have the outsider appeal of a newcomer. But few politicians with her degree of experience and pragmatism do. She is arguably the best-qualified presidential candidate for two decades.

Nonetheless, the schism in US society runs deep, and will not be healed by one election. The situation is most acute for the Republican Party, which faces an existential moment. Nobody knows what Trump’s followers will do next. America is fertile territory for conspiracy theories, and Trump is fanning the flames with allegations that the election is rigged. But his rebuke extends to the entire political system, which can be fairly accused of promoting decades of policies that put wealthy power brokers first. Cynicism is palpable on both sides of the spectrum, and the political machine built by Clinton and her coterie of advisers is ill-suited to salve these wounds.

Trump is the product of a social phenomenon that cannot be ignored. He has tapped into a much larger undercurrent of legitimate anger that is fuelling political upheaval in many countries. The Netherlands has Geert Wilders. Hungary has Viktor Orbán. France has Marine Le Pen, a more politically astute version of her father, Jean-Marie. The xenophobic and populist message spouted by such politicians is ages old and has secured the rise of countless tyrants throughout history. Most recently, hostility towards immigrants contributed to the United Kingdom’s decision to leave the European Union.

This hostility is rooted in anxiety: about cultural disruption, job and financial security and a sense that political systems are being exploited and run for the benefit of somebody else. It’s true that, for decades, Western leaders have promoted free trade and globalization as an end goal, and businesses have gradually shifted their resources around the world to gain efficiencies and bolster profits. In parallel, the rise of mechanization and robotics has reduced the need for people working in factories and on farms. As a result, millions have lost jobs in industrialized countries. And all the time, billions in the

developing world continue to struggle in poverty, often rocked by political instability and outright war.

“The schism in US society runs deep, and will not be healed by one election.”

Whatever the cause, extreme and visible inequality is a recipe for widespread political instability, and that is in nobody’s interest — including that of the global elite. This is a

central challenge for politicians today, and researchers must play their part. Many people have benefited from globalization and modern technology, and not just the rich. But too many have lost out. Unquestionably, it is time to reassess national and global economic policies with an eye towards equity and fairness. This does not mean closing borders, raising protective tariffs and putting a damper on technological development. But, clearly, politicians around the world need more and better information about how current economic policies affect people, both at home and abroad. They also need solutions.

Such questions are particularly salient when it comes to climate policy — which helps to explain why global warming is one of the few scientific issues to receive any attention at all in the US election. Clean energy represents an enormous economic opportunity, but it also poses a threat to entrenched economic interests and, in many cases, jobs.

Clinton has proposed a US\$30-billion plan to help communities that depend on coal to make the transition to a clean-energy economy. That won’t be easy, but it’s the right idea. Trump has promised to focus on fossil-fuel development and to pull out of the Paris climate treaty. Sadly, this is one issue on which his views align with Republican orthodoxy.

Indeed, the party’s official 2016 platform writes off the Intergovernmental Panel on Climate Change as a “political mechanism”, and says the modern environmental agenda is based on “shoddy science” and “scare tactics”. And as discussed in a News story on page 300, the House ‘science’ panel has become little more than a partisan attack dog.

Although both parties have become more extreme over the past two decades, conservatives have turned their backs on mainstream science to an unprecedented degree. If there is any good news, it’s that everybody now recognizes that the Republican Party has a problem. A new generation of conservative leaders will need to set a fresh course. In the meantime, Clinton must take the reins. ■

Look harder

A neuroscience initiative aims to end the invisibility of female scientists at conferences.

Relatively few women make it to top academic positions in science — and there begins the vicious circle of invisibility. Women aren't available as mentors for aspiring young scientists. They aren't there when journalists call for someone to provide a quick scientific opinion. And they are apparently not thought of when conference organizers put together lists of speakers to invite to meetings, says a group of frustrated neuroscientists trying to do something practical about the problem.

Fed up with attending meetings where most invited speakers are men, even when there are plenty of competent women to choose from, the group has created BiasWatchNeuro to bring a more systematic approach to monitoring and challenging the gender balance of academic conferences. Have a look at it: it's an eye-opener.

As successful neuroscientists themselves, the women (and a few men) behind the name-and-shame initiative know about bias-free sampling. They would like to see gender parity on speaker lists, to counter some of the many biases that hold women back. But they lobby most insistently for the minimum decency: that the percentage of women invited to speak at a particular meeting is at least equal to the base rate of women in its field.

They have worked out the base rate for neuroscience as a whole — 24% — from looking at the proportion of women in the faculties of top US universities. They use other information sources to work out the base rate for each subdiscipline — sometimes by looking at attendance lists of important meetings, more often by turning to the US National Institutes of Health (NIH) list of investigator-initiated grants, which can be searched with keywords, and simply counting up the number of female and male grant-winners. Particular subdisciplines may have other ways of working out the base rate.

Since starting in August last year, they have analysed more than

90 conferences. Two meetings last month show what makes the group angry. One was on memory mechanisms in health and disease, a subject that the NIH grant-winner list suggests has a base rate of 42% women. It mustered only 2 female invited speakers in a line-up of 17 — just 12%. The other was on tools and protocols for handling big neuroscience data, a subject in computational neuroscience, which has a low base rate of just 17–20%. The organizers managed to find no women at all to include among the 14 invited speakers.

"In our scientific society, women tend to be invisible."

Why does this happen? It is almost certainly not down to a conscious desire to exclude women. But we all unthinkingly develop biases that are shaped by the society we operate in. In our scientific society, women tend to be invisible. It's that vicious circle. Can initiatives like BiasWatchNeuro help to end it? Simply bringing the issue into open discussion in such clearly scientific terms helps a lot. The prestigious US Computational and Systems Neuroscience meeting Cosyne used to be male-dominated but, thanks to vocal complaints in the past few years, its gender ratio of invited speakers is now routinely above the field's base rate. It is one of the shining examples on BiasWatchNeuro. Its equivalent in Europe, the Bernstein Conferences, has been exposed: last year, it mustered only one female invited speaker. Whether because it felt shamed, or because BiasWatchNeuro has given women confidence to insist, it has 42% female speakers this year — well over the field's base rate.

Conference organizers should not feel that they have done their duty if they invite a top woman scientist who declines. The most successful women in science get inundated with invitations, but there will always be other successful women to choose from, and identifying them has been made easy. Anne's List (created by computational neuroscientist Anne Churchland at Cold Spring Harbor Laboratory in New York) groups female neuroscientists easily into topic and seniority level. In Europe, AcademiaNet identifies women across scientific disciplines.

The creators of BiasWatchNeuro chose the name — even though the simpler BiasWatch.com domain was available — because they hope that other scientists will get together to organize BiasWatchAnother-discipline.com. *Nature* urges you to do so. Female scientists, you have nothing to lose but your invisibility. ■

Sharp practice

Monkeys can make tool-like objects, but that doesn't mean they know what they're doing.

Technology is often a tale of seamless acquisition and refinement of skills — from rocks banged together, and bows and arrows, to steam engines and integrated circuits. But the appearance of artefacts is a different thing from their makers' intentions — if any.

As researchers show in a study published online in *Nature* this week (T. Proffitt *et al.* *Nature* <http://dx.doi.org/nature20012>; 2016), capuchin monkeys (*Sapajus libidinosus*) from the Serra da Capivara National Park in Brazil smash rocks in a way that produces sharp-edged stone flakes; were these flakes associated with an early Stone Age site, they might be regarded as intentionally produced. Indeed, progress in Stone Age technology is sometimes measured in terms of an increase in the number of sharp edges that can be coaxed from a given amount of raw material. This, of course, presupposes that producing flakes is, in fact, the intention. But capuchins, having created stone flakes, let them lie.

Why the monkeys go to all that effort remains a mystery. However, the researchers observed that about half of the monkeys sniffed or licked the broken surfaces afterwards, suggesting that they break rocks to extract mineral supplements in a conveniently powdered form.

Other monkeys bash rocks together, but the capuchins are the only wild, non-human primates known to do so with the seeming intention of breaking them. Chimps sometimes break rocks by mistake, but even when taught to bang rocks together with intent, bonobos don't create anything that resembles what is found in the hominin record.

Recognizing the earliest stone tools for what they are is not always easy, but certain features mark artefacts as the product of intent. These include the 'conchoidal' flaking that leaves a distinctive percussion mark; the production of several flakes from a single core, and the use of specific patterns of flake removal. Such features distinguish artefacts from geofacts — that is, rocks broken by natural processes, rather than objects made by non-human animals — but they say little or nothing about how the artefacts might have been used. As the capuchin example shows, the intent of the makers of the earliest artefacts can be hard to discern. The producers of the earliest stone tools to be generally recognized as such (S. Harmand *et al.* *Nature* **521**, 310–315; 2015) lived 3.3 million years ago, and were very different from modern humans.

The capuchin study should also dampen ideas that the human hand, with its precision grip, together with advanced hand-eye coordination, must necessarily have been evolutionary products or prerequisites of technology. Capuchins break rocks without the benefit of either.

In the end, the activity of banging rocks together should be seen as precisely that, and not as the first, proleptic step towards the stars. The ape-man at the start of 2001: *A Space Odyssey* that throws a bone in the air that becomes a space station was, after all, a modern human in a gorilla suit. ■

DANIEL THOMPSON



Program good ethics into artificial intelligence

Concerns that artificial intelligence will pose a danger if it develops consciousness are misplaced, says **Jim Davies**.

What is it that makes us worry about artificial intelligence (AI)? The White House is the latest to weigh in on the possible threats posed by clever machines in a report last week. As two of those involved write in a Comment piece on page 311, scientific and political focus on extreme future risks can distract us from problems that already exist.

Part of the reason for this concentration on severe, existential threats from AI comes from misplaced attention on the possibility that such technology could develop consciousness. Recent headlines suggest that respected thinkers such as Bill Gates and Stephen Hawking are concerned about machines becoming self-aware. At some point, a piece of software will 'wake up', prioritize its desires above ours and threaten humanity's existence.

But, when we worry about AI, machine consciousness is not as important as people think. In fact, careful reading of the warnings from Gates, Hawking and others show that they never actually mention consciousness. Furthermore, the fear of self-awareness distorts public debate. AI becomes defined as dangerous or not purely on the basis of whether it is conscious or not. We must realize that stopping an AI from developing consciousness is not the same as stopping it from developing the capacity to cause harm.

Where did this concern of machine consciousness come from? It seems mainly a worry of laypeople and journalists. Search for news articles about AI threats, and it's almost always the journalist who mentions consciousness. Although we do lots of things unconsciously, such as perceiving visual scenes and constructing the sentences we say, people seem to associate complicated plans with deliberate, conscious thought. It seems inconceivable to do something as complex as taking over the world without consciously thinking about it. So it could be that people have a hard time imagining that AI could pose an existential threat unless it also had conscious thought.

Some researchers argue that consciousness is an important part of human cognition (although they don't agree on what its functions are), and some counter that it serves no function at all. But even if consciousness is vitally important for human intelligence, it is unclear whether it's also important for any conceivable intelligence, such as one programmed into computers. We just don't know enough about the role of consciousness — be it in humans, animals or software — to know whether it's necessary for complex thought.

It might be that consciousness, or our perception of it, would naturally come with superintelligence. That is, the way we would judge something as conscious or not would be based on our interactions with it. A superintelligent AI would be able to talk to us, create

computer-generated faces that react with emotional expressions just like somebody you're talking to on Skype, and so on. It could easily have all of the outward signs of consciousness. It might also be that development of a general AI would be impossible without consciousness.

(It's worth noting that a conscious superintelligent AI might actually be less dangerous than a non-conscious one, because, at least in humans, one process that puts the brakes on immoral behaviour is 'affective empathy': the emotional contagion that makes a person feel what they perceive another to be feeling. Maybe conscious AIs would care about us more than unconscious ones would.)

Either way, we must remember that AI could be smart enough to pose a real threat even without consciousness. Our world already has plenty of examples of dangerous processes that are completely unconscious. Viruses do not have any consciousness, nor do they have intelligence. And some would argue that they aren't even alive.

In his book *Superintelligence* (Oxford University Press, 2014), the Oxford researcher Nick Bostrom describes many examples of how an AI could be dangerous. One is an AI whose main ambition is to create more and more paper clips. With advanced intelligence and no other values, it might proceed to seek control of world resources in pursuit of this goal, and humanity be damned. Another scenario is an AI asked to calculate the infinite digits of pi that uses up all of Earth's matter as computing resources. Perhaps an AI built with more laudable goals, such as decreasing suffering, would try to eliminate

humanity for the good of the rest of life on Earth. These hypothetical runaway processes are dangerous not because they are conscious, but because they are built without subtle and complex ethics.

Rather than obsess about consciousness in AI, we should put more effort into programming goals, values and ethical codes. A global race is under way to develop AI. And there is a chance that the first superintelligent AI will be the only one we ever make. This is because once it appears — conscious or not — it can improve itself and start changing the world according to its own values.

Once built, it would be difficult to control. So, one safety precaution would be to fund a project to make sure the first superintelligent AI is friendly, beating any malicious AI to the finish line. With a well-funded body of ethics-minded programmers and researchers, we might get lucky. ■

Jim Davies is associate professor at the Institute of Cognitive Science at Carleton University in Ottawa, Canada.
e-mail: jim@jimdavies.org

WE SHOULD PUT
MORE EFFORT
INTO PROGRAMMING
GOALS,
VALUES
AND
ETHICAL
CODES.

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

Two stars have three disks

Young stars are surrounded by a rotating disk of gas and dust, from which planets are born — but astronomers have discovered that one pair of young stars orbiting around each other has three disks, not just two.

Christian Brinch at the University of Copenhagen and his colleagues used the Atacama Large Millimeter/submillimeter Array in Chile to view a system of two stars roughly 120 parsecs (391 light years) from Earth that are each surrounded by a disk. But the authors also noticed a third, larger disk surrounding the entire system. None of the disks are aligned with each other or with the orbit of the stars themselves.

This wild misalignment suggests that the stars formed from a turbulent cloud of gas, or that a third star was recently flung out of the system.

Astrophys. J. 830, L16 (2016)

DRUG DISCOVERY

Bacteria in humans yield drug

A small molecule produced by bacteria living naturally in people can help to combat a pathogen that is resistant to many antibiotics.

Sean Brady at the Rockefeller University in New York City and his colleagues analysed the genomes of the human microbiota to identify genes predicted to encode molecules with antibiotic properties. They then synthesized these molecules and measured their antibacterial effects. One, humimycin A, was active against a strain of methicillin-resistant *Staphylococcus*

aureus (MRSA) collected from patients. MRSA-infected mice treated with humimycin A and dicloxacillin, a commercially available antibiotic, all remained alive 48 hours after infection. By contrast, at least half of the animals died after treatment with either drug alone.

Improved bioinformatic and chemical-synthesis techniques could lead to the discovery of more compounds with therapeutic potential

from the microbial world, the authors suggest.

Nature Chem. Biol. <http://dx.doi.org/10.1038/nchembio.2207> (2016)

CLIMATE CHANGE

Wildfires burn more US forest

Climate change resulting from human activities has nearly doubled the area burned by forest fires in the western

United States over the past three decades.

John Abatzoglou at the University of Idaho in Moscow and Park Williams at Columbia University in Palisades, New York, used a climate model and data on the dryness of forested areas since 1979 to assess the contribution that climate change has made to wildfires. They found that warming temperatures made the forests drier, increasing fire risk, and expanded the area burned in



NEUROTECHNOLOGY

Paralysed man with implant feels touch

A brain implant that is wired to a robotic arm has allowed a paralysed man to feel touch on the arm's fingers.

Robert Gaunt at the University of Pittsburgh in Pennsylvania and his colleagues placed electrodes in the brain of Nathan Copeland (pictured), whose legs and lower arms were paralysed 12 years ago. They positioned the electrodes in the somatosensory cortex — the brain region that receives sensory information from the body — and an area of the motor

cortex that controls hand and arm movement. The implanted electrodes are connected by wire to a computer and robotic arm. When sensors on the fingers of the robotic arm were touched, Copeland could tell which fingers were being stimulated — and sometimes which regions of those fingers.

Putting the electrodes in different parts of the brain, or implanting more of them, could increase the sensitivity of the robotic hand.

Sci. Transl. Med. 8, 361ra141 (2016)

SYLVAIN ALEM

the western part of the United States between 1984 and 2015 by about 4.2 million hectares.

Climate change also accounted for about half the increase in both the length of the fire season and the number of days with a high risk of fire. *Proc. Natl Acad. Sci. USA* <http://doi.org/brsj> (2016)

CHEMISTRY

Meteorite makes good catalyst

An iron-based mineral from a meteorite can catalyse a chemical reaction that splits water into oxygen and hydrogen, which can be used as fuel.

Some naturally occurring metallic minerals are known to have catalytic activity. Kevin Sivula and his colleagues at the Swiss Federal Institute of Technology in Lausanne studied pieces of the Namibian Gibeon meteorite, which was identified in the nineteenth century. They tested how efficiently the mineral could catalyse the oxidation of water, and found that it performed as well as synthetic iron–nickel catalysts and remained stable for 1,000 hours.

The catalytic performance emerged only after about 10 hours of operation, when a layer containing concentrated nickel, iron and cobalt with a unique 3D structure formed at the material's surface. Natural materials could inspire the creation of new kinds of catalyst, the authors suggest. *Energy Environ. Sci.* <http://doi.org/brsp> (2016)

ANIMAL COGNITION

Bees learn and 'teach' others

Bumblebees can learn to manipulate objects — and can pass their knowledge on to other bees.

Lars Chittka at Queen Mary University of London and his colleagues presented bumblebees (*Bombus terrestris*) with a disc that had been filled

with sugar water and placed under a transparent sheet of Plexiglas. To get at the disc, the bees had to pull on a string attached to it (**pictured**). Just 2 bees out of almost 300 worked out how to do this on their own; most needed stepwise training, after which more than 80% of bees were successful.

When untrained bees watched other bees getting the sugar water, they were able to learn the trick. Seeding untrained colonies with a single trained 'demonstrator' and then pairing bees from the colony with the disc apparatus eventually resulted in roughly half of the foragers learning the task. None of the foragers in the control colonies could pull the disc out.

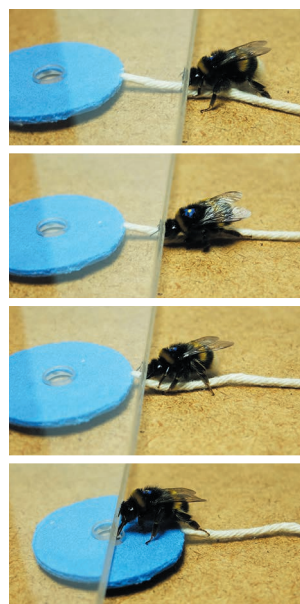
PLoS Biol. 14, e1002564 (2016)

NEUROSCIENCE

Why mole rats don't feel the heat

A gene variant could explain why naked mole rats are impervious to certain types of pain that most mammals experience when exposed to heat.

In the nervous system, a peptide called nerve growth factor (NGF) mediates hypersensitivity to pain caused by heat. Gary Lewin at the Max Delbrück Center for Molecular Medicine in Berlin and his colleagues found that in naked mole rats (*Heterocephalus glaber*; **pictured**), the chemical sequence of the NGF receptor differs from that of other vertebrates by a small handful of amino acids. As a result of these changes, the receptor fails to boost the sensitivity of another protein, TRPV1, which produces a painful, burning sensation when activated.



The authors speculate that defects in NGF signalling could also explain why, during the course of development, naked mole rats lose certain nerve fibres that conduct burning pain. This could be an adaptation to a life spent underground, where temperatures have been fairly constant for millions of years. *Cell Rep.* 17, 748–758 (2016)

PLANT BIOLOGY

RNA spray fights fungus

Spraying leaves from barley plants with a liquid containing long RNA molecules helps them to fend off fungal infection.

A mechanism called RNA interference (RNAi) uses double-stranded RNA molecules to shut down the expression of specific genes. Karl-Heinz Kogel of the Justus Liebig University in Giessen, Germany, and his colleagues used RNAi to silence three genes that fungi require to make ergosterol, a compound needed for fungal growth.

The team found that when the RNA is sprayed directly onto barley leaves, it is taken up by the fungal pathogen *Fusarium graminearum* and inhibits its growth in those leaves. Even unsprayed

leaf parts are protected from the fungus, because the RNA molecules are absorbed and transported by the leaves before being taken up by the pathogen.

The approach could open the door to a new generation of fungicides, the authors note. *PLoS Pathog.* 12, e1005901 (2016)

ELECTRONICS

Shortest transistor made

Researchers have built a transistor with a 'gate' just one nanometre long — one-fifth of the smallest length thought to be possible in silicon transistors.

The semiconductor industry is reaching the limits of its capacity to shrink silicon-based transistors. Graphene and other '2D' materials are promising replacements for silicon because they are one atom thick and have good electronic properties. Ali Javey at the University of California, Berkeley, and his collaborators have now demonstrated a transistor made of the 2D semiconductor molybdenum disulfide. In their device, a carbon nanotube laid underneath a flake of the material acts as the gate, switching the current off when a voltage is applied to it. The nanotube's one-nanometre width makes it the shortest transistor gate ever built, Javey says.

Science 354, 99–102 (2016)

CLARIFICATION

The Research Highlight 'Greenland ice loss underestimated' (*Nature* 537, 588; 2016) said that the sea-level rise from Greenland ice loss was 44% more than previous estimates. This is 44% more than some estimates. Others have given similar or greater values.

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

LORNA ELLEN FAULKES/RATS/SYSTEMS

SEVEN DAYS

The news in brief

EVENTS

All eyes on Mars

Planetary scientists are expecting the first successful landing of a European spacecraft on Mars. As *Nature* went to press, the Schiaparelli lander — part of the ExoMars joint mission with the Russian Space Agency, Roscosmos — was expected to touch down on the red planet on 19 October. The craft, which launched from Kazakhstan in March on a Russian rocket and separated from its mothership on 16 October, is intended to demonstrate landing technology, but it will also study dust storms that are expected to rage on Mars. The mission's second component, an orbiter, will begin orbiting Mars on the same day and will analyse the composition of the planet's thin atmosphere next year. See go.nature.com/2eduxjh for more.

POLICY

Green millions

The Green Climate Fund, the United Nations' financial mechanism for helping developing countries to deal with climate change, approved US\$745 million in funding proposals on 14 October. The money will go towards 10 new projects involving 27 nations. The Fund, which six years after it was launched has not yet disbursed any money, has now earmarked a total of \$1.17 billion for developing countries. At its meeting in Songdo, South Korea, the Fund's governing board also selected Howard Bamsey, former director-general of the Global Green Growth Institute and Australia's special envoy on climate change, as new executive director. Bamsey will replace H  la Cheikhrouhou, who has taken over as Tunisia's

minister for energy, mining and renewables.

Big climate win

Almost 200 nations have agreed to substantially curb their emissions of chemicals used in refrigeration and air conditioning that act as potent greenhouse gases in the atmosphere. An expansion of the Montreal Protocol, signed on 15 October at a United Nations meeting in Kigali, Rwanda, aims to reduce projected emissions of hydrofluorocarbons (HFCs) by almost 90% over the course of the twenty-first century. The protocol was created in 1987 to halt the destruction of Earth's protective ozone

layer. If left unchecked, heat-trapping HFCs, which have since replaced ozone-depleting chemicals, might have contributed up to 0.5  C of warming by the end of the century (see go.nature.com/2doehrn).

Moonshot report

The US Cancer Moonshot Task Force released a report on 17 October detailing its accomplishments and goals. The task force, which is led by US Vice President Joseph Biden, continued its call for data sharing, increased clinical trial participation and molecular tumour profiling. The moonshot aims to double the pace of cancer research.

Obama's Mars plan

US President Barack Obama reiterated his goals to send NASA astronauts to Mars in the 2030s. In an 11 October op-ed piece for CNN and at a conference in Pittsburgh, Pennsylvania, Obama said the space agency would work with private companies to develop deep-space habitats for astronauts. This includes asking companies for ideas about attaching privately built modules for living and working to the International Space Station. But with Obama leaving office in three months, the direction of NASA is up to the next president and Congress, so the goals remain uncertain at best.



MUHAID SAFODIEN/AFP/GETTY

Protests at South African universities

Student protests over tuition fees continue to disrupt teaching and academic life at South African universities. Violent clashes between students and police have been raging on campuses for several weeks despite calls from university officials to save the academic year from breakdown. Last week, protesters threw petrol bombs at buildings in the University

of KwaZulu-Natal in Durban, where a library was torched last month. At the University of Cape Town, teaching resumed on 17 October. But other campuses, including the University of the Western Cape and the Cape Peninsula University of Technology, both in Cape Town, remain closed (pictured is Vaal University of Technology near Johannesburg).

PEOPLE

Comet hunter dies

Klim Churyumov, co-discoverer of the rubber-duck-shaped comet studied by the European Space Agency's Rosetta mission, has died aged 79. Working with fellow astronomer Svetlana Gerasimenko, the Ukrainian spotted the comet using a Maksutov telescope in 1969. The space agency selected the body, known as 67P/Churyumov–Gerasimenko, as Rosetta's target in 2003, and Churyumov followed the mission closely. He lived to see its finale, a crash-landing of the mothership on the comet on 30 September. As well as being an accomplished astronomer who co-discovered a second comet in 1986, Churyumov was an avid popularizer of science and published a series of poetry books for children.

Next UN chief

Former prime minister of Portugal António Guterres (pictured, left) will be the next secretary-general of the United Nations, taking the helm after Ban Ki-moon (pictured, right) steps down on 31 December. He was appointed by the General Assembly in New York City on 13 October. Guterres, 67, studied engineering and physics in Lisbon, and had



a brief career in academia before going into politics. He was high commissioner of the UN's refugee agency for ten years until 2015, and said that alleviating the suffering of vulnerable people, and gender equality, would be key priorities for his five-year tenure as secretary-general.

RESEARCH

Galaxy glut

With the help of tens of thousands of citizen scientists from around the world, astronomers on 12 October released two data sets on the shapes of some 168,000 galaxies. The catalogues are part of the Galaxy Zoo project, which began in 2007 and has enlisted volunteers to classify nearly 1 million galaxies from the Sloan Digital Sky Survey. The latest projects (described

in two papers at <https://arxiv.org/abs/1610.03070>; 2016 and <https://arxiv.org/abs/1610.03068>; 2016) include galaxies that are farther away, with images from the Hubble Space Telescope that show galaxies up to 3.6 billion parsecs (12 billion light years) away. The results could help astrophysicists understand how galaxies have evolved.

AI manifesto

Artificial intelligence (AI) and machine learning hold significant potential for innovation and economic growth, a White House report published on 12 October concludes. Calling for government and private sector investment in research and development, the report says that regulations and standards must keep pace with the conceivable benefits of using

COMING UP

24–26 OCTOBER

Bill Gates and Richard Branson join 1,000 scientists from around the world for the Grand Challenges conference in London to share ideas on topics from crop research to menstrual hygiene.

go.nature.com/2e75xb3

2–4 NOVEMBER

The Africa Renewable Energy Forum meets in Marrakesh, Morocco, ahead of the COP22 climate meeting.

africa-renewable-energy-forum.com

AI technology in finance, health care, aviation and self-driving cars. Impacts on the economy and workforce must be scrutinized because automation in industry might particularly affect low-wage jobs, the report argues.

FACILITIES

Weighing neutrinos

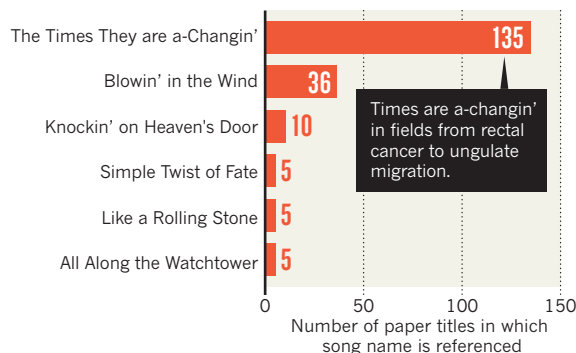
The Karlsruhe Tritium Neutrino (KATRIN) experiment in Germany, which is designed to establish the elusive mass of neutrino particles, was switched on for the first time on 14 October. Neutrinos are known to have non-zero masses, but the actual values of those masses have been difficult to measure. KATRIN will weigh the extremely light particles indirectly by measuring the energies of electrons shooting out from the nuclear decay of tritium, an isotope of hydrogen. Researchers have now started beaming electrons inside the 70-metre-long, €60-million (US\$66-million) machine, and plan to begin the tritium experiment — expected to last five years — in late 2017.

TREND WATCH

Bob Dylan, who on 13 October became a Nobel laureate in literature, might be scientists' favourite musician. A 2015 analysis found that Dylan's song names were mentioned in at least 213 article titles (C. Gornitzki, A. Larsson and B. Fadeel *Br. Med. J.* 351, h6505; 2015); numerous fields were found to be “a-changin’”. The authors concluded that Dylan's respect for the medical profession — as evidenced by his lyric “I wish I'd have been a doctor” — is reciprocated.

BOB DYLAN IN THE SCIENTIFIC LITERATURE

Songs by Bob Dylan, a newly minted Nobel literature laureate, are referenced in at least 213 paper titles in PubMed. These are the six songs that are most often mentioned (and mangled).



Based on analysis in C. Gornitzki et al. *BMJ* 351, h6505 (2015).

NATURE.COM

For daily news updates see:

www.nature.com/news

NEWS IN FOCUS

US ELECTION The scientists who support Donald Trump in the presidential race **p.298**

POLITICS US House science committee flexes its muscle **p.300**



INFORMATICS Effort to wrangle geoscience data stalls **p.303**

FORECASTS Prediction markets harness the wisdom of crowds **p.308**

KIKE CALVO/VISUAL & WRITTEN SL/ALAMY



The Arecibo Observatory has one of the world's biggest single-dish telescopes.

ASTRONOMY

Arecibo Observatory hit with discrimination lawsuit

Two former workers say that they were treated unfairly on the basis of age and disability.

BY TRACI WATSON

Two former researchers at the troubled Arecibo Observatory in Puerto Rico have filed a lawsuit claiming that illegal discrimination and retaliation led to their dismissal.

James Richardson and Elizabeth Sternke are suing the Universities Space Research Association (USRA), which oversees radio astronomy and planetary science at Arecibo, and the observatory's deputy director, Joan Schmelz — a

prominent advocate for women in astronomy.

Richardson and Sternke, a married couple in their mid-50s, allege that Schmelz discriminated against them because of their age and because Richardson is legally blind. Sternke revealed in November 2015 that she planned to file a complaint with the US Equal Opportunity Commission (EEOC), which investigates workplace bias; soon afterwards, USRA announced that her contract job with Arecibo's education programme would end early. Richardson filed his own EEOC complaint, and in

April 2016, USRA terminated his employment as a staff scientist.

The EEOC ultimately found evidence of discrimination against Sternke and Richardson, and that the pair were terminated in retaliation for their complaints. In their lawsuit, filed on 4 October in the US District Court in Puerto Rico, Richardson and Sternke are seeking more than US\$20 million in back pay and damages.

Schmelz says that she cannot comment on the lawsuit, and she declined to answer *Nature's* questions. But USRA, her co-defendant and ►

► employer, “firmly denies these allegations and plans to vigorously defend this matter”, it said in a statement to *Nature*.

The legal challenge comes as the 53-year-old observatory battles to survive. Its single-dish radio telescope, one of the world’s biggest, is still in high demand. But the US National Science Foundation (NSF), which provides roughly two-thirds of the observatory’s \$12 million funding, is facing a budget crunch. The agency is now conducting an environmental review of major changes to the site, a possible prelude to mothballing or even demolishing the facility. Its decision on Arecibo’s fate is expected in 2017.

Some Arecibo supporters worry that the lawsuit could nudge the observatory closer to the edge. “With all those budget difficulties they’re having now, getting bad press is not going to be good for them,” says Alan Harris of the planetary-science consulting firm MoreData! in La Cañada, California.

LEADERSHIP CHANGES

USRA hired Richardson in 2014 as a scientist with Arecibo’s planetary radar group, which observes potentially dangerous asteroids and other Solar System bodies. He did not follow the typical academic path: according to Richardson’s website, he worked as a nuclear engineer before being blinded in a chemical accident and retraining as a planetary scientist. Sternke, a sociologist, began working at Arecibo on a short-term contract in 2015.

According to EEOC determinations issued in June, Sternke and Richardson’s work initially drew no complaints from management. After Richardson’s boss, the head of planetary radar, announced his resignation in early 2015, Richardson sought the job.

Several months later, Schmelz came to

Arecibo. From the start, the lawsuit says, she “ignored and/or chose to avoid all contact” with Richardson, assigned duties to younger colleagues rather than to him, and “marginalized and ostracized” Richardson and Sternke.

The EEOC report also says that USRA altered the description of the job Richardson wanted “to make it more suitable for another internal candidate to qualify”. USRA subsequently promoted an Arecibo staffer in his 30s.

Sternke submitted her resignation in November. She later told USRA that she planned to file a complaint with the EEOC, the agency’s report says, and her employment was terminated on 4 December, eight days before her scheduled last day.

The lawsuit alleges that in December 2015, officials from the USRA human-resources department accused Richardson of “angry behavior, bullying, and prejudices”. His employment was terminated in April 2016, after USRA determined that he failed to meet the terms of its “Performance Improvement Plan”. (Richardson disagrees with that assessment.)

In its report on Richardson’s case, the EEOC said that Schmelz “made direct discriminatory age based comments”, writing in her own performance evaluation that she had recruited “a set of effective young leaders”.

The EEOC also found that Richardson was “disciplined and terminated from his employment” on the basis of his age and disability, and in retaliation for his association with Sternke and for filing an EEOC charge. In a separate report, the agency found that USRA terminated Sternke’s employment “due to her age (over 50) and in retaliation for complaining about illegal discrimination”.

The EEOC suggested that USRA pay Richardson \$400,000 in damages, plus back pay, and give Sternke \$200,000. But settlement talks with the EEOC failed, and in late July the agency notified Richardson and Sternke that they had 90 days to file suit.

SADNESS AND SURPRISE

Richardson’s former colleagues say that he is not a bully. “I never heard him raise his voice, let alone get angry,” says Phillip Nicholson, an astronomer at Cornell University in Ithaca, New York, where Richardson did research.

Richardson’s postdoctoral supervisor at Cornell, astronomer Joseph Veverka, describes him as courteous and kind, if demanding. “If anyone asked Jim to do something which he did not consider completely scientifically proper, he would strongly object.”

Former Arecibo director Robert Kerr says that his USRA colleagues — including Schmelz — displayed “the utmost professionalism”. “Joan was no different from the rest,” he adds.

Meg Urry, an astrophysicist at Yale University in New Haven, Connecticut, notes that Schmelz is a tireless advocate for the right of female astronomers to work without harassment. “She’s devoted a lot of time to justice,” says Urry, the past president of the American Astronomical Society. In one notable case, Schmelz helped to bring harassment complaints against astronomer Geoff Marcy; after the University of California, Berkeley, found that Marcy had violated its policies on harassment, he retired in late 2015.

The district court in Puerto Rico has not yet scheduled a hearing on the Arecibo lawsuit. In the meantime, Nicholson is struggling to make sense of the situation, given what he knows of the parties. “Nothing seems to ring true to the character of the people,” he says. ■

US ELECTION

Scientists who back Trump

Science policy fades into background for many who support the Republican candidate.

BY SARA REARDON

Kaylee, a structural biologist at Yale University in New Haven, Connecticut, stays quiet when her colleagues talk about politics and religion. As a Catholic with conservative tendencies, she feels that her beliefs are unwelcome in academic institutions, where liberal views often prevail. The strain is particularly acute this year: Kaylee favours Donald Trump for US president.

Trump, a Republican, has a run a brash, often divisive, campaign that has prompted

some leading members of his own party to disavow him. He has drawn criticism for his treatment of women, his pledge to block Muslim immigration to the United States, and his plan to build a wall along the US–Mexico border. Still, Kaylee says, “I am 100% certain I will not vote for Hillary Clinton,” Trump’s Democratic opponent, despite her fears that supporting Trump could harm her job prospects. (For this reason, Kaylee — a postdoc — asked *Nature* to refer to her by a pseudonym.)

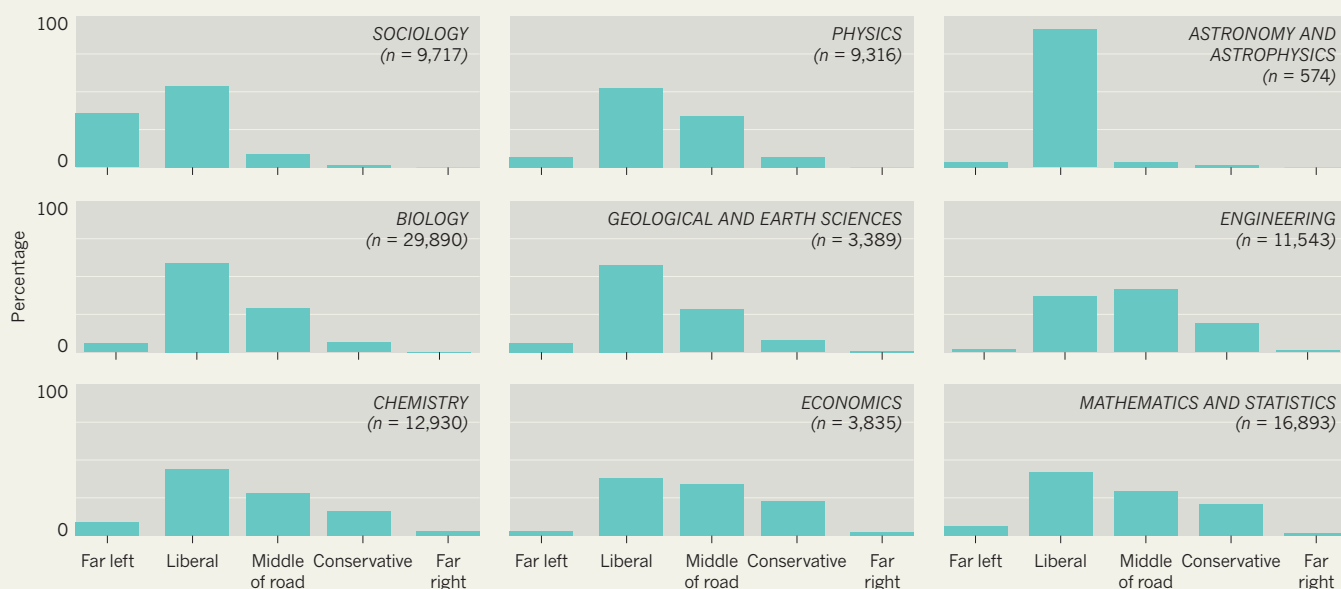
Her fears do not surprise Neil Gross, a sociologist at Colby College in Waterville, Maine.

Surveys have shown that conservative faculty members are a minority in US universities, although the proportion varies by field (see ‘Field reports’). “My sense is that the candidacy of Donald Trump has really intensified disputes that were there already in academic life,” Gross says. “If Republicans in academia and science felt uncomfortable before, I think the candidacy of Mr Trump has made them all the more uncomfortable.”

Many of the researchers interviewed for this article say that Trump and Clinton’s positions on science have not influenced their

FIELD REPORTS

US academics tend towards liberal political views, but dipping into data from a 2013–14 survey of university faculty members reveals differences between individual disciplines.



vote — in part because the candidates have largely ignored these issues on the campaign trail. “We’re living in a two-dimensional world: how much do you like each candidate, and how much do you hate each candidate?” says Stanley Young, assistant director for bioinformatics at the National Institute of Statistical Sciences in Research Triangle Park, North Carolina, who backs Trump. “The popular impression I get is Clinton would go forward with business as usual and Trump is likely to upset things a bit. There’s a lot that could be improved in science.”

David Deming, a geophysicist at the University of Oklahoma in Norman, doesn’t think it matters whether Trump and Clinton have much personal knowledge

of science. “Trump said he’d appoint good people and I believe him,” says Deming, who has written newspaper opinion pieces in support of Trump.

Other scientists who plan to vote for the Republican say they have been let down by US President Barack Obama, and think that Clinton — another Democrat — would bring more of the same. To them, Trump represents change. “The current status quo seems like it’s not working for a lot of Americans,” says one Trump-supporting chemist at the University of Pittsburgh in Pennsylvania, who asked for anonymity. “I’m hopeful for a modest improvement, and that’s about as much as I can hope.”

William Briggs, a statistician at Cornell University in Ithaca, New York, likes the fact that Trump has not emphasized science. “The federal government has become far too involved in setting the scientific agenda,” says Briggs, who argues that Obama has misused science in politically charged debates over climate change and energy policy. “I think Hillary would worsen that.”

Kaylee, who disagrees with Trump’s views on women and minorities, says that her desire for a more conservative Supreme Court is driving her vote. With the next president likely to nominate at least one Supreme Court justice — a lifetime appointment — she sees Trump as a tool to move the court’s ideological balance to the right. Otherwise, Kaylee would vote for a ‘write-in candidate’ who won’t appear on the presidential ballot: her lab’s principal investigator, who has given her a safe space to express conservative views.

But not everyone is so lucky. And as the 8 November election nears, talk of the hard-fought presidential race grows trickier to escape. Some scientists who support Trump worry that political discussions in the lab will not only harm their careers in the long term, but also hinder current collaborations with colleagues, and waste time.

“I’ve avoided discussions among my real-life peers for a while,” says the anonymous chemist at the University of Pittsburgh, who prefers to talk about politics online. “A lot of people, if they’re not willing to come out in favour of Hillary, will give the third-party dodge.” ■

“I am 100% certain I will not vote for Hillary Clinton.”



Like his opponent, Donald Trump has not emphasized science issues during his campaign.



Representative Lamar Smith (Republican, Texas) has aggressively probed how science is done.

POLITICS

House science panel flexes its muscle

Chairman Lamar Smith has turned once-placid panel into investigative powerhouse.

BY JEFF TOLLEFSON

Lamar Smith has made his mark on science. As chairman of the US House Committee on Science, Space, and Technology, the Texas Republican has launched dozens of investigations into alleged wrongdoing by scientists, environmental groups and government officials. And he shows no signs of slowing down.

Since 2013, Smith has probed everything from individual National Science Foundation grants to government air-quality regulations — issuing an unprecedented 24 subpoenas along the way. And although the Republican presidential candidate, Donald Trump, is foundering in the polls, the party is poised to retain control of the House of Representatives in the 8 November election. That means that Smith is likely to remain at the helm of the science committee for at least two more years.

Whatever the future brings, one thing is clear: the panel has shed its long-standing reputation as a bastion of collegial, bipartisan debate. “This committee is a microcosm of Congress as a whole,” says David Goldston, who served as chief of staff to former chairman Sherwood Boehlert (Republican, New York) from 2001 to 2006. “Things have gotten ever

more polarized, and at some point, the science committee wasn’t going to be an exception.”

Although he won the chairmanship four years ago, Smith didn’t shift his investigations into high gear until 2015. That’s when the committee voted along party lines to grant him unilateral authority to issue subpoenas — a powerful tool to compel witness testimony or access to sensitive documents.

At that point, the panel had not issued a subpoena since the early 1990s, when it probed safety and pollution violations at a US government nuclear-weapons facility in Colorado. But Smith has taken liberal advantage of his new authority, aided by an influx of staff recruited from another House committee that specializes in investigations.

“It’s just been one case after another,” says Representative Eddie Bernice Johnson of Texas, the highest-ranking Democrat on the panel. “Members of the committee seem to be somewhat perplexed that we got to this point.”

But the panel’s Republican staff says that such complaints are sour grapes, and note

“Members of the committee seem to be somewhat perplexed that we got to this point.”

that Smith has sought a role for Democrats in several probes. “There is a knee-jerk reaction — no matter what investigation it is — to criticize the majority,” says Mark Marin, the Republican staff director for two of the science panel’s subcommittees.

GETTING WARMER

Many of Smith’s highest-profile investigations have targeted the science underlying global warming and policies intended to reduce greenhouse-gas emissions. Last year, he sought to compel the US National Oceanic and Atmospheric Administration (NOAA) to release documents related to a study that disputed the idea of a global warming ‘pause’ around the turn of the twenty-first century (T. R. Karl *et al.* *Science* **348**, 1469–1472; 2015). Smith suggested that NOAA scientist Thomas Karl had altered data to advance an “extreme climate-change agenda”, which drew a sharp rebuke from the agency and science organizations.

In July, Smith subpoenaed the attorneys-general of New York and Massachusetts over their push to determine whether oil giant Exxon Mobil misled investors about the financial liabilities posed by climate change. Smith, who has accused the state officials of trying to stifle legitimate scientific debate, is seeking documents and other communications regarding the states’ probe. He has also issued subpoenas to eight environmental groups that have sought to determine whether fossil-fuel companies knowingly spread false information about climate science.

Smith declined *Nature*’s request for an interview. In a statement, he said that his investigations are meant to defend the “freedom of scientific inquiry” — and the interests of taxpayers. “I plan on carrying out my responsibility to protect the First Amendment rights of scientists and continuing our constitutional oversight responsibility,” he wrote.

Thus far, many of Smith’s subpoenas have come to naught. In some cases, such as the Exxon probe, the committee’s targets have argued that Smith has overstepped his constitutional authority. The Massachusetts and New York attorneys-general and the eight environmental groups have declined to comply with the science panel’s subpoenas. In the case of the NOAA climate study, the agency briefed the committee and provided some documents, but withheld internal communications between its scientists.

To enforce a disputed subpoena, the full House would need to vote in favour of holding the recipient in contempt of Congress by the end of the year — an unlikely scenario.

Nonetheless, Smith can start afresh when the new Congress convenes in early 2017. Marin hints that this is likely for the Exxon probe, at least. “The chairman is interested in continuing this investigation until he gets what he is looking for,” he says. ■

BILL CLARK/CQ ROLL CALL/GETTY

REPRODUCTIVE BIOLOGY

Mouse eggs made in the lab

First eggs created wholly in a dish raise call for debate over technology's use in humans.

BY DAVID CYRANOSKI

In a tour de force of reproductive biology, scientists in Japan have transformed mouse skin cells into eggs in a dish, and used those eggs to birth fertile pups. The report marks the first creation of mouse eggs entirely outside the animal. Researchers hope the process could be adapted to produce lab-grown human eggs too.

Katsuhiko Hayashi, a reproductive biologist at Kyushu University in Fukuoka, led the group that announced the breakthrough on 17 October in *Nature* (O. Hikabe *et al.* *Nature* <http://doi.org/brxt>; 2016). In 2012, when at the University of Kyoto, he and stem-cell biologist Mitinori Saitou reported taking skin cells down the pathway towards eggs: reprogramming them to embryonic-like stem cells and then into primordial germ cells (PGCs). These early cells emerge as an embryo develops, and later give rise to sperm or eggs. But to get the PGCs to form mature eggs, the researchers

had to transfer them into the ovaries of living mice. The next advance came in July 2016, when a team led by Yayoi Obata at the Tokyo University of Agriculture reported transforming PGCs extracted from mouse fetuses into oocytes (egg cells) without using a live mammal. Working with Obata, Hayashi and Saitou have now completed the progression: from skin cells to functional eggs in a dish. With the use of *in vitro* fertilization techniques, 26 healthy pups were born, and some of them have given birth to offspring.

"This is truly amazing," says Jacob Hanna, a stem-cell biologist at the Weizmann Institute of Science in Rehovot, Israel. "To be able to make robust and functional mouse oocytes over and over again entirely in a dish, and see the entire process without the 'black box' of having to do any of the steps in host animals, is most exciting." The procedure is technically challenging, Hayashi says, but different groups in his lab have reproduced it. Although the team did not

need to implant PGCs into living mice, they did have to add cells from ovaries of other mouse fetuses, effectively creating an ovary-like support in which the eggs could grow.

Hayashi says the work will help him to study egg development; he is not trying to make functional human eggs in the lab. But he suspects that others will try. "I do not think it is going to prove much more complex," says Hanna. Hayashi thinks that "oocyte-like" human eggs might be produced within ten years, but doubts that they will be of sufficient quality for fertility treatments. In his study, only 3.5% of the early embryos created from artificial eggs gave rise to pups, compared with 60% of eggs that were matured inside a mouse.

Debate over the ethics of the technology should begin now, says Azim Surani, a pioneer in the field at the University of Cambridge, UK. "This is the right time to involve the public in these discussions, long before the procedure becomes feasible in humans," he says. ■

PHILANTHROPY

Science group guides Silicon Valley philanthropists

Uncertain government funding drives efforts to increase private support for research.

BY ERIKA CHECK HAYDEN

Poring over a 2014 list of the top 50 philanthropists in the United States, physicist Marc Kastner noticed that 16 were based in California, compared with just 6 in New York, Connecticut and New Jersey combined.

"That made it pretty clear where I should be," says Kastner, who established the offices of the Science Philanthropy Alliance in Palo Alto, California, when he was named the organization's first president in February 2015. The alliance is made up of philanthropic organizations that encourage funding for basic research and advise other philanthropists — especially new ones — on how to go about it.

That bet paid off last month, when Facebook founder Mark Zuckerberg and physician and educator Priscilla Chan announced that their Chan Zuckerberg Initiative would spend at least US\$3 billion on medical research over the next decade. In remarks describing the initiative's plan to eliminate, manage or prevent all major disease by 2100, Zuckerberg urged fellow philanthropists to seek advice from Kastner.

"This is a milestone for the alliance, in the sense that its goal is to try to increase the funding for basic research across many fields, not just in biology," says Robert Tjian, former president of the Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland.

BASIC SUPPORT

The HHMI was one of 6 founders of the alliance, which now counts 15 members. They include heavy hitters such as the Palo Alto-based Gordon and Betty Moore Foundation, the Kavli Foundation in Oxnard, California, and the London-based Wellcome Trust. Smaller up-and-coming members include the Eric and

Wendy Schmidt Fund for Strategic Innovation in Palo Alto and the Heising-Simons Foundation, based in Los Altos, California.

The alliance was formed in 2013 after a budget crunch hammered US government funding for research. The group focuses on boosting private giving to basic research, as pressure intensifies on agencies such as the US National Institutes of Health to fund more applied and translational research.

Kastner says he was drawn to his current job after watching young physicists struggling to start their careers at the Massachusetts Institute of Technology in Cambridge, where he worked for more than 40 years. "There was a need to support basic research that wasn't being met, and I thought I could help by having philanthropists do some of that," he says.

The physicist is poised to make a major impact on the shape of US philanthropy as its centre of gravity shifts west, away from New York, where philanthropic families such

as the Rockefellers and the Carnegies based their foundations in the early twentieth century. West-coast philanthropy is anchored by established players such as the Bill and Melinda Gates Foundation in Seattle, Washington. But new ones are joining at a rapid clip, including the Parker Foundation in San Francisco, California, founded by Napster's Sean Parker. He has spent upwards of \$280 million on cancer, allergy and malaria research.

A GUIDING HAND

Chan and Zuckerberg's announcement showcases how the Science Philanthropy Alliance works. Initially, the pair considered funding translational research. But after conversations over many months with the alliance and other advisers, they decided to concentrate on fundamental research, especially the building of scientific tools and technologies. The alliance provided the couple with connections to scientists, as well as advice on defining basic research and how to structure and run a scientific advisory board.

"Philanthropists don't want us to tell them what kind of research to support, but they want to make sure they're doing it in the best way possible, and that's what we want, too," says Kastner.

The alliance also convenes private meetings at which members discuss questions such as how to measure the impact of philanthropic spending, and whether science funding is best spent on people, places or projects, says Harvey Fineberg, president of the Gordon and Betty Moore Foundation.

Crucially, the alliance does all its work on a confidential basis, so Kastner won't say what's coming next. But he promises that it will be significant: "We're talking to people who have the potential of making very big contributions to basic science." ■



Marc Kastner (right) and his group advise donors.

DOMINICK REUTER


**MORE
ONLINE**

TOP NEWS



Nations agree to ban refrigerant gases to reduce warming
go.nature.com/2doehr

MORE NEWS

- Observable Universe has 2 trillion galaxies
go.nature.com/2emdwgt
- Unusual object seen in outer Solar System
go.nature.com/2eifcj5
- Bet on human lifespan to pay off in 2150
go.nature.com/2elqyko

VIDEO OF THE WEEK



How the European Space Agency plans to land a probe on Mars
go.nature.com/2eduxjh

ESA/ATG MEDIALAB

GLEB GARANICH/REUTERS



EarthCube is meant to give scientists access to data from stored geological samples, such as these ice cores.

INFORMATICS

Geoscience data project struggles

Five years in, the US EarthCube programme has found it hard to deliver on its promises.

BY ALEXANDRA WITZE

A US National Science Foundation programme to help geoscientists handle ever-increasing amounts of data is facing a mid-life crisis.

Called EarthCube, the five-year-old geoinformatics effort was conceived as a game-changer: it would put obscure data online, link and enrich existing databases across disciplines and develop software tools for scientists. But in March, an external advisory panel warned that EarthCube still lacked a clear definition and might not be sustainable. The project's leaders have been working to overhaul it, and by the end of this month they aim to release what could be a make-or-break plan for the US\$13-million programme.

"We need to make changes in order to show progress and to pull the community back in," says Kerstin Lehnert, a geologist at the Lamont-Doherty Earth Observatory in Palisades, New York, who heads EarthCube's leadership council. "We have to make this work, now."

EarthCube is the broadest effort yet to bring US geoscience into the modern data era. Some fields are more up to date than others,

says Catherine Constable, a palaeomagnetist at the Scripps Institution of Oceanography in La Jolla, California, who co-led the recent review. She notes that US seismologists have compiled their earthquake recordings into well-managed products that are easy to access, whereas geochemists tend to collect individual measurements on individual rock samples that stay mostly in their labs. "There are areas of geoscience with data hidden away in people's drawers," Constable says.

EarthCube is supposed to change that by providing tools to give scientists access to rich troves of otherwise-hidden data. "It's the embodiment of a vision that many of us have had for many years," says Dawn Wright, a marine geologist and chief scientist at Esri, a data-mapping company in Redlands, California. "But there are a lot of growing pains."

The programme's early efforts focused on integrating data across disciplines. The iSamples initiative built a rich database describing physical samples, from sea-floor cores and river-water samples to fossils, so that researchers could track studies being done with each specimen. The GeoLink project developed a way to mine information from published

resources, including research papers, field reports and laboratory analyses.

But many say that EarthCube has yet to deliver on its early promise. Charles Connor, a volcanologist at the University of South Florida in Tampa, received EarthCube funding to develop software tools for tracking ash and other volcanic hazards, but says his colleagues are happy with the popular VHub.org research platform. And an ambitious computer-science effort called CINERGI, which was meant to compile data sets and documentation across many fields, has yet to move beyond pilot demonstrations.

EarthCube may be a victim of its ambition, trying to do too much for too many people. The March review noted that five years in, the programme still had poorly defined goals. "There wasn't a uniform vision of what they wanted to do," Constable says. That may be because EarthCube has been a grass-roots effort, led by a small group of passionate volunteers. The review recommended reorganizing with fewer leaders to set clear priorities for the entire research community.

Lehnert says that EarthCube's leaders are taking the criticism to heart. They pulled together a rapid-response team that has been outlining ways to restructure the programme, including introducing metrics — such as the number of software downloads — to assess its performance. There will be more details in the draft plan set for release later this month, she says. "People need to be patient. EarthCube will help tremendously, but development does take time."

Eva Zanzerkia, the EarthCube programme manager at the NSF, notes that technology develops rapidly, whereas the social aspects of sharing data take a while to develop. "This contrast is something EarthCube grapples with regularly," she says.

There are some promising EarthCube tools on the horizon. Constable points to the Geoscience Papers of the Future project, which encourages researchers to publish their findings with all the metadata and tools needed for others to replicate their studies. It launched in 2015 and is run by early-career scientists who are pushing for transparency in research.

Five years from now, EarthCube should be nimble enough to quickly give scientists the rich, interlinked data they need, says Lisa Park Boush, a palaeobiologist and palaeoclimatologist at the University of Connecticut in Storrs who helped to set up the programme. "We'll be able to do in an easy way what would take days, weeks or months now," she says. ■

CORRECTION

The News story 'Where Nobel winners start' (*Nature* **538**, 152; 2016) wrongly said that the study assessed only four categories of the Nobel awards, in fact it looked at all six.



THE POLLING CRISIS: **HOW TO TELL WHAT PEOPLE REALLY THINK**

This year's US presidential election is the toughest test yet for political polls as experts struggle to keep up with changing demographics and technology. *BY RAMIN SKIBBA*

Hillary Clinton is heading for a landslide victory over Donald Trump. But wait. Trump is pulling ahead and could take the White House. No, Clinton has a clear lead and is gaining ground. Nearly every day, a new poll comes out touting a different result, leaving voters wondering what to believe.

The results of recent elections give even more reason for scepticism.

In 2013, the Liberal Party of Canada confounded expectations when it won the provincial elections in British Columbia. The following year, polls overestimated support for Democrats in the US congressional elections. And this year, some pollsters underestimated Britons' support for leaving the European Union in the Brexit referendum. These blunders have led some political commentators to say that polls are headed for the graveyard.

"It's harder and harder to find people willing to pay for any polls, given their poor performance this year and last year. They're heavily discredited in the UK," says Stephen Fisher, a political sociologist at the University of Oxford.

As the US presidential election approaches, pollsters are scrambling to improve their methods and avoid another embarrassing mistake. Their job is getting harder. Until as recently as ten years ago, polling organizations were able to tap into public opinion simply by calling people at home. But large segments of the population in developed countries have given up their landlines for mobile phones. That is making them more difficult for pollsters to reach because people will often not answer calls from unfamiliar numbers.

So the pollsters are fighting back. They are fine-tuning their efforts in reaching mobile phones, using statistical tools to correct for biases and turning to online surveys. The increasing number of online polls has prompted the formation of polling aggregates, such as FiveThirtyEight, RealClearPolitics and Huffington Post, which combine and average the results to develop more nuanced forecasts.

"Polling's going through a series of transitions. It's more difficult to do now," says Cliff Zukin, a political scientist at Rutgers University in New Brunswick, New Jersey. "The paradigm we've used since the 1960s has broken down and we're evolving a new one to replace it — but we're not there yet."

The ingredients of an accurate poll are fairly simple, but they can be hard to find, and everyone uses a different recipe to pull them all together. Start by recruiting a large group of people — preferably more than 1,000. The sample should be split evenly between women and men. And it should reflect the population's mix in terms of race, education, income and geographical distribution, to represent these groups' different views and voting behaviours. Once the data are in hand, pollsters analyse the gaps in their sample and weight the results to account for groups that are under-represented.

"Polling is an art, but it's largely a scientific endeavour," says Michael Link, president and chief executive of Abt SRBI polling firm in New York City and former president of the American Association for Public Opinion Research.

It's also a process that is conducted behind closed doors. Polls are run by a mix of companies and academic groups, but they are generally commissioned by news organizations and political groups. As a result, pollsters rarely share the details of their techniques. "There's a lot of people who make a living doing this, and whose reputations are set on it," says Jill Darling, survey director at the University of Southern California's Center for Economic and Social Research in Los Angeles.

CHANGING TIMES

The data-gathering part of polling used to be relatively easy in developed countries. Pollsters simply called people at home — at first, by hand, and later with automatic diallers in the United States. But landlines are quickly going the way of the telegraph (see 'The line on voters'). In 2008, more than eight in every ten US households had landlines; by 2015, that number had dropped to five and it continues to decline. In the United Kingdom, more people have landlines but the fraction is dropping. As of this year, 53% of them claim that they never or rarely use them.

The mobile revolution has hit pollsters hard in the United States because federal regulations require that mobile phones be called manually. And people often do not answer calls to their mobiles when an unfamiliar number pops up. In 1997, pollsters could get a response rate of 36% but that has dropped to just 10% or less now. As a result, pollsters are struggling to reach as many people, and costs are going up: each mobile-phone interview costs about twice as much as a landline one. There is also a 'non-response bias', because people who respond to pollsters' calls sometimes do not reflect a representative sample, says Frederick Conrad, head of the Program in Survey Methodology at the University of Michigan in Ann Arbor.

Despite the expense and difficulty of calling people, this method still produces the most accurate results, says Courtney Kennedy, director of survey research at the Pew Research Center in Washington DC. US pollsters now call mobile phones for more than half of their samples, and that fraction will probably rise as more and more people ditch their landlines.

Pollsters are also grappling with another major problem — predicting who will vote. That is likely to be unusually difficult in the United States this year because many voters aren't enamoured of the leading candidates, who have historically low approval ratings.

US national elections typically have turnouts of 40–55%, lower than most other developed countries, according to the Organisation for Economic Co-operation and Development. In the United Kingdom, by contrast, 60–70% of the eligible population usually votes. Richer, older, better-educated people, and those who voted in the previous election, are more likely to vote, but this varies with each election.

Pollsters typically base their estimates of turnout on their own proprietary mix of factors such as respondents' voting history, whether they're registered with a political party, their engagement with politics, whether they say they're planning to vote, as well as demographics and socioeconomic factors. "Likely voter" modelling is notoriously the secret-sauce aspect of polling," says Kennedy.

It's also one of the most difficult parts of accurate polling. In the

2014 mid-term US election, most pollsters failed in their forecasts of Democratic voting. Turnout was just 36% — a record low in the past 70 years — which disproportionately depressed votes for Democratic candidates.

In the 2015 UK general election, most major pollsters, including ICM Unlimited and YouGov, underestimated the turnout of older, Conservative Party voters, according to an inquiry published in March by the British Polling Council and Market Research Society¹. The inquiry also found that pollsters have systematic biases in their samples. They tend to have too many Labour supporters at the expense of Conservative ones. They had applied weighting and adjustment procedures to the raw data, but this has not mitigated the bias problem. Another source of error identified in the report is "herding" — when pollsters consciously or unconsciously adjust their polls so that their results seem similar to those released earlier, causing the polls to converge.

The bias in favour of left-leaning parties is not unique to the United Kingdom. The inquiry analysed more than 30,000 polls from 45 countries and found a similar, although smaller, bias. The report did not give an explanation for why, but some pollsters in the United States and Britain attribute the trend to inaccurate predictions of who will turn up to vote.

"POLLING IS AN ART, BUT IT'S LARGELY A SCIENTIFIC ENDEAVOUR"

In the case of the United Kingdom, the panel recommended that pollsters work to obtain more representative samples and to investigate better ways to weight them.

Pollsters are also trying to improve their accuracy by changing how they model likely voters. In the past, they treated their sample in a binary fashion: determining how many would turn out on election day and how many would stay home. Now they tend to assign a probability to whether someone will vote.

More transparency could help. Pollsters in the United Kingdom share their methodologies with the British Polling Council, which aided the recent investigation and has led to fruitful debates about ways to improve accuracy, says Fisher, who participated in the inquiry.

IN DATA WE TRUST

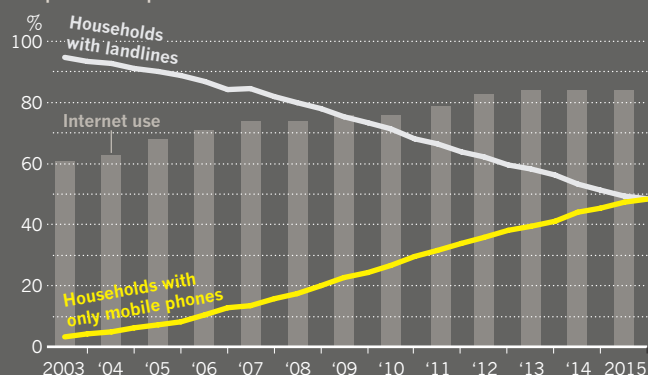
Even if polling organizations manage to collect a representative sample, they can't always trust the responses that people give them. One of the starkest examples in the United States came in the 1982 election for California's governor. Los Angeles Mayor Tom Bradley, an African American, was consistently leading in the polls but lost the election by a narrow margin. Afterwards, pollsters suggested that the discrepancy arose because some voters might not have wanted to admit that they would not support an African American candidate. This is now known as the 'Bradley effect'.

A variation on this is the 'shy Tory effect', named after Conservative-leaning voters in the United Kingdom who hide their views or misreport their intentions to pollsters. That makes some experts wonder whether a shy Trump effect might come into play in the forthcoming US election — in which a fraction of voters are embarrassed about or reluctant to admit their support for Trump or opposition to Clinton. But most major pollsters doubt that this will be a major factor

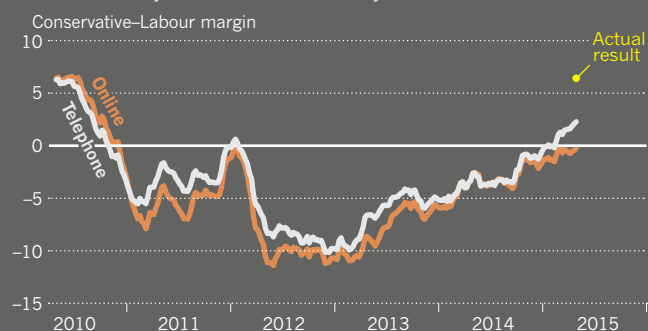
THE LINE ON VOTERS

Pollsters have had trouble getting an accurate read on voters' intentions in some recent elections.

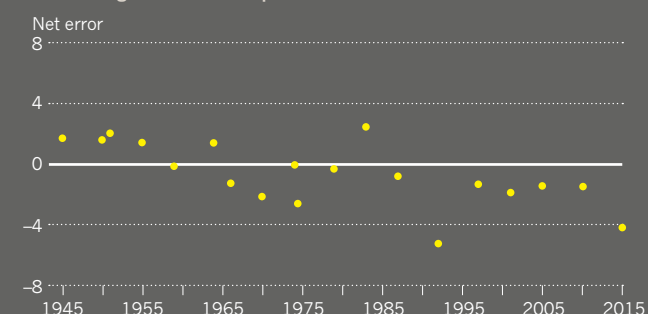
The switch from landlines to mobile phones has made it harder for some polling organizations in the United States to get large, representative samples. Online polls have become more common as Internet use has risen.



Polls in the United Kingdom failed to predict the strong showing of the Conservative Party in the 2015 Parliamentary elections.



Polls in the United Kingdom have been underestimating the share of votes going to the Conservative Party for decades, but the error in 2015 was larger than for most previous elections.



because polls before the Republican primary elections gauged support for Trump accurately and he has performed similarly in online polls and in ones that use live interviews.

Advanced technology may allow pollsters to get a better read on voters' true feelings. Online polls, for instance, allow people to respond at their convenience and state their intentions without fear of judgement from a live interviewer. They also make it easy to collect thousands of responses in a short time and at a lower cost: about US\$30,000 for a 12-minute survey as opposed to more than US\$70,000 for a similar telephone one, says Chris Jackson, vice-president at Ipsos Public Affairs, a global market-research and polling firm in Washington DC.

But online polls have challenges, too. They typically recruit by advertising on popular websites, so people choose whether to participate, and that means that there might be a built-in bias in their

samples. Pollsters don't exactly know who is missing from the poll, and it's harder to estimate the reliability of the final poll numbers.

Some pollsters have begun experimenting with polls conducted through text messages. As with online polls, people can choose to respond whenever they want and avoid talking to a person. Michael Schober, a psychologist at The New School for Social Research in New York City, and his colleagues tested the differences between live and text interviews². "The lack of time pressure and social pressure of texting leads people to disclose more information and be more honest," he says.

Another approach is to assemble a panel of people to survey repeatedly. The most prominent is a University of Southern California Dornsife/Los Angeles Times Presidential Election tracking poll that launched in July. These pollsters randomly selected people on the basis of information from the US Postal Service and contacted them by mail, recruiting 3,000 people to participate each week in their online surveys. Unlike other polls, they need not continually recruit new respondents, and their response rate is at least 15% — higher than for telephone polls. The pollsters have enough data to know the demographics of their sample very well and can have confidence in their trends, says Darling, who leads the survey.

However, if their sample turns out to be biased, then all polls for the duration of the sample will be biased. This may be the case with this year's poll, which leans slightly towards Trump, according to the aggregator FiveThirtyEight.

To reduce the risk of bias, researchers are experimenting with a new type of poll. Andrew Gelman, a statistician and political scientist at Columbia University in New York City, and his colleagues have collected a very large set of people and divided them up into tens of thousands of demographic categories. The researchers tested this extreme categorization method on polling data from the 2012 US presidential election, showing that it produced accurate forecasts of state-level results by using highly tuned weights to correct for the non-representative sample³. However, this sophisticated method takes much more time and requires more detailed data than are usually gathered.

It could be a glimpse of the future, however. 'Big data' are where more accurate results will come from, says Joe Twyman, head of political and social research for Europe, Middle East and Africa at YouGov. "It will be about linking a respondent's voting data with Internet usage, other survey data, and demographic information, creating a much richer picture of that person, which will allow for more accurate granulations of predictions," he says. Pollsters would use this information to assess who is likely to vote and to analyse the survey results — for example, by determining which issues most concern different voters.

The low cost of Internet polling has triggered a surge in the number of polls of varying quality, making it hard for journalists, policymakers and others to separate the wheat from the chaff. Poll aggregators attempt to weight polls on the basis of the past reliability, but that doesn't guarantee future success, especially if low-quality and short-lived polling outfits are included in the mix.

Contrary to bold claims of the death of polls, practitioners say that they are merely going through a transition. But pollsters do recognize that some of the barriers are insurmountable. As election seasons lengthen and people find more reasons to survey public opinion, the number of polls will continue to rise. Pollsters recognize that they can only ask so much of people, says Gelman. "There's a non-renewable resource of public trust." ■

Ramin Skibba is an intern with *Nature* in Washington DC.

1. Sturgis, P. et al. *Report of the Inquiry into the 2015 British General Election Opinion Polls* (Market Research Society and British Polling Council, 2016); available at <http://eprints.ncrm.ac.uk/3789>
2. Schober, M. F. et al. *PLoS ONE* **10**, e0128337 (2015).
3. Wang, W., Rothschild, D., Goel, S. & Gelman, A. *Int. J. Forecast.* **31**, 980–991 (2015).



MARKET *FORECASTS*

**Prediction markets
can be uncannily
accurate — sometimes.
Scientists have begun
to understand why
they work, and how
they can fail.**

BY ADAM MANN

It was a great way to mix science with gambling, says Anna Dreber. The year was 2012, and an international group of psychologists had just launched the ‘Reproducibility Project’ — an effort to repeat dozens of psychology experiments to see which held up¹. “So we thought it would be fantastic to bet on the outcome,” says Dreber, who leads a team of behavioural economists at the Stockholm School of Economics.

In particular, her team wanted to see whether scientists could make good use of prediction markets: mini Wall Streets in which participants buy and sell ‘shares’ in a future event at a price that reflects their collective wisdom about the chance of the event happening. As a control, Dreber and her colleagues first asked

a group of psychologists to estimate the odds of replication for each study on the project’s list. Then the researchers set up a prediction market for each study, and gave the same psychologists US\$100 apiece to invest.

When the Reproducibility Project revealed last year that it had been able to replicate fewer than half of the studies examined², Dreber found that her experts hadn’t done much better than chance with their individual predictions. But working collectively through the markets, they had correctly guessed the outcome 71% of the time³.

Experiments such as this are a testament to the power of prediction markets to turn individuals’ guesses into forecasts of sometimes startling accuracy. That uncanny ability ensures

ILLUSTRATION BY SÉBASTIEN THIBAUT

that during every US presidential election, voters avidly follow the standings for their favoured candidates on exchanges such as Betfair and the Iowa Electronic Markets (IEM). But prediction markets are increasingly being used to make forecasts of all kinds, on everything from the outcomes of sporting events to the results of business decisions. Advocates maintain that they allow people to aggregate information without the biases that plague traditional forecasting methods, such as polls or expert analysis (see page 304).

In science, applications might include giving agencies impartial guidance on the proposals that are most worth funding, helping panels to find a consensus in climate science and other fields or, as Dreber showed, giving researchers a fast and low-cost way to identify the studies that might face problems with replication.

But sceptics point out that prediction markets are far from infallible. "There is a viewpoint among some people that once you set up a market this magic will happen and you'll get a great prediction no matter what," says economist Eric Zitzewitz at Dartmouth College in Hanover, New Hampshire. That is not the case: determining the best designs for prediction markets, as well as their limitations, is an area of active research.

Nevertheless, prediction-market supporters argue that even imperfect forecasts can be helpful. "Hearing there's an 80 or 90% chance of rain will make me take an umbrella," says Anthony Aguirre, a physicist at the University of California, Santa Cruz. "I think there's a big space between being able to time travel and physically see what will happen, and then throwing up your hands and saying it's totally unpredictable."

THE MAGIC OF GAMBLING

People have been betting on future events for as long as they have played sports and raced horses. But in the latter half of the nineteenth century, US efforts to set betting odds through marketplace supply and demand became centralized on Wall Street, where wealthy New York City businessmen and entertainers were using informal markets to bet on US elections as far back as 1868. These political betting pools lasted into the 1930s, when they fell victim to factors such as stricter gambling laws and the rise of professional polling. But while they lasted they had an impressive success rate, correctly picking the winners of 11 out of 15 presidential races, and correctly identifying that the remaining 4 contests would have extremely tight margins.

The prediction-market idea was revived by the spread of the Internet, which dramatically lowered the entry barriers for creating and participating in prediction markets. In 1988, the University of Iowa's Tippie College of Business

launched the not-for-profit IEM as a network-based teaching and research tool; ahead of the 8 November presidential election that year, they set up a market to predict the fraction of votes that would go to each of the presidential candidates (see 'How a market predicts'). The fractions changed daily as traders interpreted fresh

participated in prediction markets that beat the firm's official projections of quarterly printer sales 75% of the time. And in September 2002, six months before the US-led invasion of Iraq, the Dublin-based betting site TradeSports.com gained international notoriety when it ran a prediction market on when Iraqi dictator Saddam Hussein would be ousted. By the time the war began in March 2003, bettors were 90% certain Hussein would be out by April and 95% sure he'd be gone by May or June. He was deposed in April.

MARKET RESEARCH

Prediction markets have also had some high-profile misfires, however — such as giving the odds of a Brexit 'stay' vote as 85% on the day of the referendum, 23 June. (UK citizens in fact narrowly voted to leave the European Union.) And prediction markets lagged well behind conventional polls in predicting that Donald Trump would become the 2016 Republican nominee for US president.

Such examples have inspired academics to probe prediction markets. Why do they work as well as they do? What are their limits, and why do their predictions sometimes fail?

Perhaps the most fundamental answer to the first question was provided in 1945 by Austrian economist Friedrich Hayek. He argued that markets in general could be viewed as mechanisms for collecting vast amounts of information held by individuals and synthesizing it into a useful data point — namely the price that people are willing to pay for goods or services.

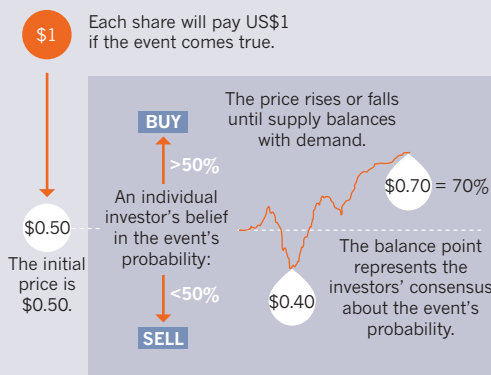
Economists theorize that prediction markets do this information gathering in two ways. The first is through 'the wisdom of crowds' — a phrase popularized by business journalist James Surowiecki in his book of that name (Doubleday, 2004). The idea is that a group of people with a sufficiently broad range of opinions can collectively be cleverer than any individual. An often-cited case is a game

in which participants are asked to estimate the number of jelly beans in a jar. Although individual guesses are unlikely to be right, the accumulated estimates tend to form a bell curve that peaks close to the actual answer. When investor Jack Treynor ran this experiment on 56 students in 1987, their mean estimate for the number of beans — 871 — was closer to the correct answer of 850 than all but one of their guesses⁵.

As Surowiecki and others have emphasized, however, crowds are wise only if they harbour a sufficient diversity of opinion. When they don't — when people's independent judgements are skewed by peer pressure, panic or even a charismatic speaker — the wisdom of crowds can easily fall prey to collective breakdowns. The housing bubble of the mid-2000s, which was a major contributor to the 2007–08 financial

HOW A MARKET PREDICTS

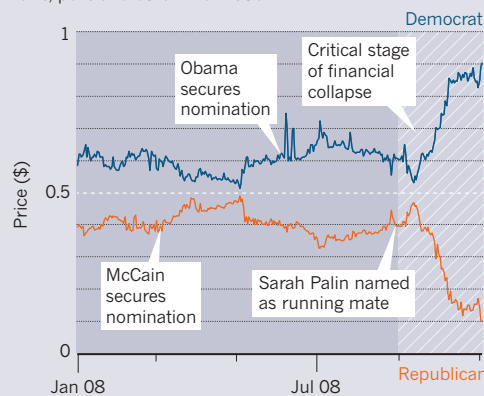
Prediction markets use investors' opinions to generate a price for 'shares' in a given event.



EXAMPLE: 2008 US PRESIDENTIAL ELECTION

Barack Obama (Democrat) versus John McCain (Republican)

The market's prediction constantly shifted to accommodate news, polls and other information.



information about polls, the economy and other issues. On the eve of the election, the market predicted that the Republican nominee, George H. W. Bush, would be victorious with 53.2% of the vote — which is exactly what he got. And in 2008, a study found that the IEM's predictions across five presidential elections were more accurate than the polls 74% of the time⁴.

The success of the IEM helped to inspire the creation of dozens of other prediction markets. In 1996, for example, the Hollywood Stock Exchange was launched to forecast opening-weekend box-office take and other movie-related outcomes; its markets correctly predicted that *Hamlet* would be a flop that year and that *Jerry Maguire* would be a hit. In the early 2000s, employees of information-technology company Hewlett-Packard

crash, was one such breakdown of judgement. But this is where the second market mechanism comes in. Sometimes called the marginal-trader hypothesis, it describes how — in theory — there will always be individuals seeking out places where the crowd is wrong. In the process, these traders will identify undervalued contracts to buy and overvalued contracts to sell, which tends to push prices back towards a sensible value. An example can be seen in the 2015 film *The Big Short*, which dramatizes the true story of a hedge fund that bet against the irrational exuberance of the US housing market and gained substantially from the crash.

its funding ended. But it helped to inspire Metaculus, a market launched in November 2015 by Aguirre and his colleague Greg Laughlin, an astrophysicist now at Yale University in New Haven, Connecticut. The site grew out of Aguirre's interest in finding 'superpredictors' — people whose forecasting skills are far above average. Metaculus asks participants to estimate the probabilities of such things as, "Will a clinical trial begin by the end of 2017 using CRISPR to genetically modify a living human?" or "Will the National Ignition Facility announce a shot at break-even fusion by the start of 2017?"

As in SciCast, Metaculus participants do

markets, however: "It is in general easier to make short-term than long-term predictions," says Aguirre. As long as prediction markets offer a way to update guesses in light of new information, proponents argue, they will do as well or better than other forecasting methods.

Scientific prediction markets also suffer more from ambiguity issues than do political or economic ones. In an election, one person is eventually declared the winner, whereas in science, resolutions are rarely so neat. But prediction-market advocates don't think that this is necessarily a cause for concern. "When someone starts to suggest a bet, people immediately start to clarify what they mean," says Hanson. Aguirre says that he and Laughlin take great pains on Metaculus to ensure that predictions are well-defined and easy to understand.

Whether prediction markets can work for science remains an open question. When Dreber's team repeated 18 economics experiments as part of a follow-up to her psychology investigation, both the prediction markets and surveys of individuals overestimated the odds of each study's reproducibility⁹. Dreber isn't sure why this happened. She points out that the psychologists in the first study were all already interested in replication — whereas the economists in the second were not involved in the reproducibility project — so they might have been better at collectively estimating reproducibility.

Prediction markets in general still need to deal with challenges such as how to limit manipulation and overcome biases. Yet conventional representative polling, which once relied on answers from phonecalls to randomly sampled landlines, is being jeopardized by the movement to mobile phones and online messaging. Because the accuracy of prediction markets is at least on par with, if not better than, polls, economist David Rothschild of Microsoft Research in New York City thinks that prediction markets are well placed to take over if polling goes into decline. "I can create a poll that can mimic everything about a prediction market," he says. "Except markets have a way of incentivizing you to come back at 2 a.m. and update your answer." ■

Adam Mann is a freelance writer in Oakland, California.

"WHEN SOMEONE STARTS TO SUGGEST A BET, PEOPLE IMMEDIATELY START TO CLARIFY WHAT THEY MEAN."

Laboratory experiments have been used to test many aspects of this theoretical framework, including how well prediction markets aggregate information under different conditions. In a 2009 experiment⁶ that was designed to mimic scientific research and publishing, researchers set up three prediction markets in which participants tried to predict which hypothesis about a fictitious biochemical pathway would end up being true.

FIELD-TESTING THE FUTURE

In one market, key pieces of information about the pathway were available to all participants; the traders quickly converged on the correct answer. In another, analogous to proprietary corporate research, information was privately held by individuals; the traders often failed to reach a consensus. And in the third, analogous to results being discovered in different labs and then published in journals, information was initially kept private and then made public. The market was able to find the right answer — but the individuals who discovered useful information first could use their private knowledge to anticipate the markets and extract a small profit.

One of the first prediction markets devoted exclusively to scientific questions grew out of a project started in 2011 by economist Robin Hanson of George Mason University in Fairfax, Virginia. Eventually known as SciCast, the project included a website where participants could wager on questions such as, "Will there be a lab-confirmed case of the coronavirus Middle East Respiratory Syndrome (MERS or MERS-COV) identified in the United States by 1 June 2014?" (There was.) SciCast's assessments were more accurate than an uninformed prediction model 85% of the time (see go.nature.com/2dm6llp).

SciCast was discontinued in 2015, when

not use actual money: players instead move a slider representing their belief in the likelihood of an answer and accrue a track record for being correct. The lack of cash bets is partly a matter of practicality, says Aguirre. "When it's 'Will Hillary win?', zillions of people will buy on that. But if it's 'Will this new paper on arXiv get more than ten citations?', you're not going to find enough people with real money to make an accurate prediction." But it's also the case that real money isn't strictly necessary for a successful prediction market: several studies^{7,8} have shown that traders can be equally well motivated by the prestige of being right.

Metaculus currently has around 2,000 active users, although its creators hope to accrue 10,000 or more. Already, the site has produced evidence that successful prediction is a skill that can be learned. The best players work out the optimal time to adjust their guesses up or down, and their performance gradually improves.

Laughlin and Aguirre suggest that Metaculus could be useful to journalists and other members of the public who want to know which questions most interest scientists. Funding agencies might similarly be attracted to its results. "Having prediction markets that are getting an even-handed assessment is potentially a way of aiding the decision for what projects are most worth funding," says Laughlin.

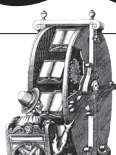
But scientific prediction markets have yet to gain much traction with researchers or the public. One important reason is that most political and business questions get clear-cut answers in relatively short time periods, and this is where prediction markets excel. But few would-be traders have the patience to endure the decades of effort, ambiguity and experimentation that are often required to answer questions in science.

This problem is hardly unique to prediction

1. Open Science Collaboration *Perspect. Psychol. Sci.* **7**, 657–660 (2012).
2. Open Science Collaboration *Science* **349**, aac4716 (2015).
3. Dreber, A. *et al. Proc. Natl Acad. Sci. USA* **112**, 15343–15347 (2015).
4. Berg, J. E., Nelson, F. D. & Rietz, T. A. *Int. J. Forecast.* **24**, 285–300 (2008).
5. Treynor, J. L. *Financ. Anal. J.* **43**, 50–53 (1987).
6. Almenberg, J., Kittlitz, K. & Pfeiffer, T. *PLoS ONE* **4**, e8500 (2009).
7. Rosenbloom, E. S. & Notz, W. *Electron. Mark.* **16**, 63–69 (2006).
8. Servan-Schreiber, E., Wolfers, J., Pennock, D. M. & Galebach, B. *Electron. Mark.* **14**, 243–251 (2004).
9. Camerer, C. F. *et al. Science* **351**, 1433–1436 (2016).

COMMENT

TREES A celebration of the poet of forest ecology, Oliver Rackham **p.314**



ENGINEERING On the many uses of rotation, from biology to business **p.315**

PALAEONTOLOGY Why are women in the field donning beards for equality? **p.316**

OBITUARY Deborah Jin, pioneer of ultracold physics, remembered **p.318**

ARMANDO L. SANCHEZ/CHICAGO TRIBUNE/TNS/GETTY



Chicago police use algorithmic systems to predict which people are most likely to be involved in a shooting, but they have proved largely ineffective.

There is a blind spot in AI research

Fears about the future impacts of artificial intelligence are distracting researchers from the real risks of deployed systems, argue **Kate Crawford** and **Ryan Calo**.

On 12 October, the White House published its report on the future of artificial intelligence (AI) — a product of four workshops held between May and July 2016 in Seattle, Pittsburgh, Washington DC and New York City (see go.nature.com/2dx8rv6).

During these events (which we helped to organize), many of the world's leading thinkers from diverse fields discussed how AI will change the way we live. Dozens of presentations revealed the promise of using progress in machine learning and other AI techniques

to perform a range of complex tasks in everyday life. These ranged from the identification of skin alterations that are indicative of early-stage cancer to the reduction of energy costs for data centres.

The workshops also highlighted a major blind spot in thinking about AI. Autonomous systems are already deployed in our most crucial social institutions, from hospitals to courtrooms. Yet there are no agreed methods to assess the sustained effects of such applications on human populations.

Recent years have brought extraordinary

advances in the technical domains of AI. Alongside such efforts, designers and researchers from a range of disciplines need to conduct what we call social-systems analyses of AI. They need to assess the impact of technologies on their social, cultural and political settings.

A social-systems approach could investigate, for instance, how the app AiCure — which tracks patients' adherence to taking prescribed medication and transmits records to physicians — is changing the doctor-patient relationship. Such an approach ►

► could also explore whether the use of historical data to predict where crimes will happen is driving overpolicing of marginalized communities. Or it could investigate why high-rolling investors are given the right to understand the financial decisions made on their behalf by humans and algorithms, whereas low-income loan seekers are often left to wonder why their requests have been rejected.

A SINGULAR PROBLEM

“People worry that computers will get too smart and take over the world, but the real problem is that they’re too stupid and they’ve already taken over the world.” This is how computer scientist Pedro Domingos sums up the issue in his 2015 book *The Master Algorithm*¹. Even the many researchers who reject the prospect of a ‘technological singularity’ — saying the field is too young — support the introduction of relatively untested AI systems into social institutions.

In part thanks to the enthusiasm of AI researchers, such systems are already being used by physicians to guide diagnoses. They are also used by law firms to advise clients on the likelihood of their winning a case, by financial institutions to help decide who should receive loans, and by employers to guide whom to hire.

Analysts are expecting the uses of AI systems in these and other contexts to soar. Current market analyses put the economic value of AI applications in the billion-dollar range (see ‘On the rise’), and IBM’s chief executive Ginni Rometty has said that she sees a US\$2-trillion opportunity in AI systems over the coming decade. Admittedly, estimates are difficult to make, in part because there is no consensus on what counts as AI.

AI will not necessarily be worse than human-operated systems at making predictions and guiding decisions. On the contrary, engineers are optimistic that AI can help to detect and reduce human bias and prejudice. But studies indicate that in some

current contexts, the downsides of AI systems disproportionately affect groups that are already disadvantaged by factors such as race, gender and socio-economic background².

In a 2013 study, for example, Google searches of first names commonly used by black people were 25% more likely to flag up advertisements for a criminal-records search than those of ‘white-identifying’ names³. In another race-related finding, a ProPublica investigation in May 2016 found that the proprietary algorithms widely used by judges to help determine the risk of reoffending are almost twice as likely to mistakenly flag black defendants than white defendants (see go.nature.com/29aznyw).

THREE TOOLS

How can such effects be avoided? So far, there have been three dominant modes of responding to concerns about the social and ethical impacts of AI systems: compliance, ‘values in design’ and thought experiments. All three are valuable. None is individually or collectively sufficient.

Deploy and comply. Most commonly, companies and others take basic steps to adhere to a set of industry best practices or legal obligations, so as to avoid government, press or other scrutiny. This approach can produce short-term benefits. Google, for example, tweaked its image-recognition algorithm in 2015 after the system mislabelled an African American couple as gorillas. The company has also proposed introducing a ‘red button’ into its AI systems that researchers could press should the system seem to be getting out of control⁴.

Similarly, Facebook made an exception to its rule of removing images of nude children from its site after the public backlash about its censorship of the Pulitzer-prizewinning photograph of a naked girl, Kim Phúc, fleeing a napalm attack in Vietnam. And just last month, several leading AI companies, including Microsoft, Amazon and IBM,

formed the Partnership on AI to try to advance public understanding and develop some shared standards.

Yet the ‘deploy and comply’ approach can be ad hoc and reactive, and industry efforts can prove inadequate if they lack sufficient critical voices and independent contributors. The new AI partnership is inviting ethicists and civil-society organizations to participate. But the concern remains that corporations are relatively free to field test their AI systems on the public without sustained research on medium- or even near-term effects.

Values in design. Thanks to pioneers in the ethical design of technology, including the influential scholars Batya Friedman and Helen Nissenbaum, researchers and firms now deploy frameworks such as value sensitive design or ‘responsible innovation’ to help them to identify likely stakeholders and their values. Focus groups or other techniques are used to establish people’s views about personal privacy, the environment and so on. The values of prospective users are then incorporated into the design of the technology, whether it is a phone app or a driverless car⁵. Developers of AI systems should draw on these important methods more.

Nevertheless, such tools often work on the assumption that the system will be built. They are less able to help designers, policymakers or society to decide whether a system should be built at all, or when a prototype is too preliminary or unreliable to be unleashed on infrastructure such as hospitals or courtrooms.

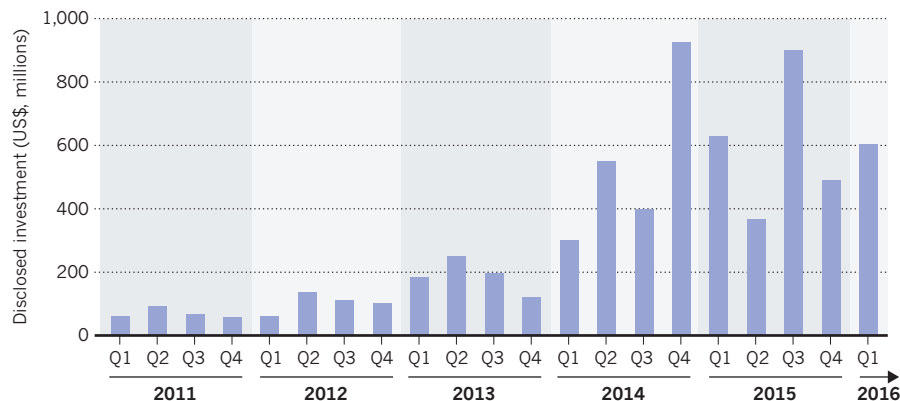
Thought experiments. In the past few years, hypothetical situations have dominated the public debate around the social impacts of AI.

The possibility that humans will create a highly intelligent system that will ultimately rule over us or even destroy us has been most discussed (see, for example, ref. 6). Also, one relevant thought experiment from 1967 — the trolley problem — has taken on new life. This scenario raises questions about responsibility and culpability. In it, a person can either let a runaway trolley car run along a track where five men are working, or pull a lever to redirect the trolley on to another track where only one person is at risk. Various commentators have applied this hypothetical scenario to self-driving cars, which they argue will have to make automated decisions that constitute ethical choices⁷.

Yet as with the robot apocalypse, the possibility of a driverless car weighing up ‘kill decisions’ presents a narrow frame for moral reasoning. The trolley problem offers little guidance on the wider social issues at hand: the value of a massive

ON THE RISE

Investment in technologies that use artificial intelligence has climbed in recent years.





People with asthma were wrongly graded as low risk by an AI system designed to predict pneumonia.

investment in autonomous cars rather than in public transport; how safe a driverless car should be before it is allowed to navigate the world (and what tools should be used to determine this); and the potential effects of autonomous vehicles on congestion, the environment or employment.

SOCIAL-SYSTEMS ANALYSIS

We believe that a fourth approach is needed. A practical and broadly applicable social-systems analysis thinks through all the possible effects of AI systems on all parties. It also engages with social impacts at every stage — conception, design, deployment and regulation.

As a first step, researchers — across a range of disciplines, government departments and industry — need to start investigating how differences in communities' access to information, wealth and basic services shape the data that AI systems train on.

Take, for example, the algorithm-generated 'heat maps' used in Chicago, Illinois, to identify people who are most likely to be involved in a shooting. A study⁸ published last month indicates that such maps are ineffective: they increase the likelihood that certain people will be targeted by the police, but do not reduce crime.

A social-systems approach would consider the social and political history of the data on which the heat maps are based. This might require consulting members of the community and weighing police data against this feedback, both positive and negative, about the neighbourhood policing. It could also mean factoring in findings by oversight committees and legal institutions. A social-systems analysis would also ask whether the risks and

rewards of the system are being applied evenly — so in this case, whether the police are using similar techniques to identify which officers are likely to engage in misconduct, say, or violence.

As another example, a 2015 study⁹ showed that a machine-learning technique used to predict which hospital patients would develop pneumonia complications worked well in most situations. But it made one serious error: it instructed doctors to send patients with asthma home even though such people are in a high-risk category. Because the hospital automatically sent patients with asthma to intensive care, these people were rarely on the 'required further care' records on which the system was trained. A social-systems analysis would look at the underlying hospital guidelines, and other factors such as insurance policies, that shape patient records⁹.

A social-systems analysis could similarly ask whether and when people affected by AI systems get to ask questions about how such systems work. Financial advisers have been historically limited in the ways they can deploy machine learning because clients expect them to unpack and explain all decisions. Yet so far, individuals who are already subjected to determinations resulting from AI have no analogous power¹⁰.

A social-systems analysis needs to draw on philosophy, law, sociology, anthropology and science-and-technology studies, among other disciplines. It must also turn to studies of how social, political and cultural values affect and are affected by technological

“Artificial intelligence presents a cultural shift as much as a technical one.”

change and scientific research. Only by asking broader questions about the impacts of AI can we generate a more holistic and integrated understanding than that obtained by analysing aspects of AI in silos such as computer science or criminology.

There are promising signs. Workshops such as the Fairness, Accountability, and Transparency in Machine Learning meeting being held in New York City next month is a good example. But funders — governments, foundations and corporations — should be investing much more in efforts that approach AI in the way we describe.

Artificial intelligence presents a cultural shift as much as a technical one. This is similar to technological inflection points of the past, such as the introduction of the printing press or the railways. Autonomous systems are changing workplaces, streets and schools. We need to ensure that those changes are beneficial, before they are built further into the infrastructure of everyday life. ■ **SEE WORLD VIEW P.291**

Kate Crawford is a principal researcher at Microsoft Research in New York City, a visiting professor at the Massachusetts Institute of Technology in Cambridge, Massachusetts, and a senior research fellow at New York University, New York, USA.

Ryan Calo is an assistant professor of law and of information science (by courtesy), and faculty co-director of the Tech Policy Lab at the University of Washington, Seattle, Washington, USA.

e-mails: kate@katecrawford.net; rcalo@uw.edu

1. Domingos, P. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (Allen Lane, 2015).
2. Barocas, S. & Selbst, A. D. *Calif. Law Rev.* **104**, 671–732 (2016).
3. Sweeney, L. *Discrimination in Online Ad Delivery* (2013); available at <http://dx.doi.org/10.2139/ssrn.2208240>
4. Armstrong, S. & Orseau, L. in *Uncertainty in Artificial Intelligence: Proceedings of the Thirty-Second Conference* (eds Ihler, A. & Janzing, D.) 557–566 (AUAI Press, 2016); available at <http://go.nature.com/2drokil>
5. Friedman, B., Kahn, P. H. & Borning, A. in *Human-Computer Interaction in Management Information Systems: Foundation* (eds Zhang, P. & Galletta, D.) 348–372 (M. E. Sharpe, 2006); available at <http://go.nature.com/2dee8om>
6. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies* (Oxford Univ. Press, 2016).
7. Lin, P. in *Autonomes Fahren: Technische, Rechtliche und Gesellschaftliche Aspekte* (eds Maurer, M., Gerdes, J. C., Lenz, B. & Winner, H.) 69–85 (Springer, 2015); available at <http://doi.org/brdw>
8. Saunders, J., Hunt, P. & Hollywood, J. S. *J. Exp. Criminol.* **12**, 347–371 (2016).
9. Caruana, R. et al. 'Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission' *Proc. 21th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* 1721–1730 (ACM, 2015).
10. Crawford, K. et al. *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term* (2016); available at <https://artificialintelligence.now.com>



Oliver Rackham analysed interactions between people and woodlands.

NATURAL HISTORY

Voices from the greenwood

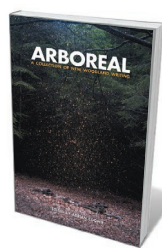
Caspar Henderson applauds a paean to the brilliant forest ecologist Oliver Rackham.

For too many years, I have been one of the rats running frequently along the motorway from Oxford through the Chiltern Hills into London. On the capital's outskirts, my bus sweeps past a scrap of land wedged between the road and a London Underground line, where an increasingly decrepit sign proclaims the future home of gleaming corporate headquarters. In reality, the site has, over time, turned from rubbish-strewn concrete to a dense young wood dominated by birch trees, some more than 8 metres tall. The last time I went by, on a bright, blustery autumn day, their tens of thousands of leaves caught the sunlight in a vision of glory.

This patch of self-willed wood might have raised a smile from botanist and historical ecologist Oliver Rackham. Over several decades until his death in 2015, Rackham probably did more than any other scientist to advance our understanding of the consequences of human interactions with woodlands and other landscapes in Britain. That contribution inspires and informs *Arboreal*, a beautiful, insightful anthology edited by nature writer Adrian Cooper. The more than 40 pieces by ecologists, educators, photographers, sculptors and writers, are highly diverse. Their common starting point is that

the perceptions, memories and imagination of individuals matter, and that without wonder and reflection, research and action are blind and blundering.

Rackham's insights remain compelling. With more than a dozen books, notably *Ancient Woodland* (Edward Arnold, 1980) and *The History of the Countryside* (J. M. Dent & Sons, 1986), he attracted and educated a large public. He showed that ancient woodland — any wood in continuous existence since at least 1600 — was frequently as much a human artefact as 'natural.' He helped to set in train significant changes to planning and conservation, fighting the Forestry Commission over the planting of uniform stands of conifers, and showing that large unmanaged populations of deer pose serious threats to woods. In his last years, he stressed the dangers of unregulated global trade in trees, a factor in the spread of pests and diseases.



Arboreal: A Collection of New Woodland Writing
EDITED BY ADRIAN COOPER
Little Toller: 2016.

His final book, *The Ash Tree* (Little Toller, 2014), explored the place of the much-loved genus *Fraxinus* in culture, and explained Chalara ash dieback, caused by the fungus *Hymenoscyphus fraxineus*, which threatens to wipe out a large proportion of this hitherto abundant species. (In 2012, dieback was reported to have affected some 90% of ash trees in Denmark, but Rackham withheld judgement about the probable scale of UK devastation.) Above all, Rackham helped to create a vision of woodlands as complex, dynamic and potentially resilient places that for thousands of years have seldom been free from human intervention, and where the stories we tell and choices we make have significant consequences.

Cooper calls this long relationship one of both estrangement and affection. Britain is among the most deforested countries in Europe: almost half the ancient woodland in the British Isles was either felled or poisoned between 1933 and 1983. Yet the country famously treasures its large ancient trees. "The majority of us do not own woodlands nor earn our living by them," Cooper writes in the introduction, "yet it seems that the trees and the woods still inhabit us."

So George Peterken, a woodland ecologist

EAST ANGLIAN DAILY TIMES

of Rackham's generation, unfolds the surprises and paradoxes of what may be Britain's most 'natural' ancient wood, at Lady Park in the Wye Valley. Gabriel Hemery, forest scientist and author of *The New Sylva* (Bloomsbury, 2014; G. Hemery *Nature* **507**, 166–167; 2014), 'looks back' from 2050 on the reforestation of Dartmoor in the face of climate change.

Poet Kathleen Jamie brings the form and sensibility of classical Chinese poetry to the woodlands around Inshriach Bothy in Scotland's Cairngorms National Park. Jay Griffiths (author of *Wild*; Hamish Hamilton, 2007) unfolds the beauty in woodland birdsong, in prose of great energy and power. Sculptor David Nash explains how he created the extraordinary *Ash Dome* at Cae'n-y-Coed in Snowdonia, Wales. And in images such as one of leaf-fall through an opening in the New Forest canopy, photographer Ellie Davies creates a sense of immanence redolent of Andrei Tarkovsky's 1975 film *The Mirror*.

Elsewhere, Tobias Jones — co-founder of the Windsor Hill Wood project for young people with mental-health issues — makes a powerful case for the therapeutic use of woodland. He explores the Japanese practice of *shinrin yoku*, or forest bathing, and its reputed beneficial effects on insomnia and anxiety. Deb Wilenski, a specialist in early-childhood education, shows how Spinney Wood in Cambridgeshire can fire the imaginations of the very young to produce maps and poems (B. Kiser *et al. Nature* **523**, 286–289; 2015). And music journalist Will Ashon recounts the social history of Epping Forest — Britain's unofficial first national park and "a Cockney Paradise".

In *I Contain Multitudes*, an account of the microbiome (Ecco, 2016; A. Woolfson *Nature* **536**, 146–147; 2016), journalist Ed Yong describes dysbiosis. This process of irreversible decline triggered by perturbation from factors including antibiotics or pollution can afflict the human gut and other complex ecosystems, such as tropical coral reefs. *Arboreal*, which itself resembles a thicket of ancient woodland — unruly and pulsing with life, full of surprises and beauty in both detail and the long view — offers consolation and counsel for those who hope to save Britain's woods from such a fate. Trees are not, to paraphrase the poet and artist William Blake, green things standing in the way. They are living communities that are part of us, as we are still, in myriad subtle ways, part of them. ■

Caspar Henderson is the author of *The Book of Barely Imagined Beings*. His *New Map of Wonders* will be published next year.
e-mail: caspar81@gmail.com

BIOMECHANICS

The wonders of whirl

John E. Moalli and Adam P. Summers relish a book on biomechanical spin, from wheels to free-falling felines.

In *Why the Wheel Is Round*, biomechanist Steven Vogel (who died last year) succeeds once again in turning engineers, biologists and the general public onto the beauty, complexity and approachability of his field. He spins an 11-part tale of circular motion that ranges from rotation in biology to rotation driven by biology. Vogel captivates with discussions of engineering feats rooted in circular motion — from plodding horses turning shallow paddle wheels to gears that drive sixteenth-century reading machines — and doesn't stint on his trademark puns and word-play. Mixing findings in his own field with those from mechanics, dynamics and historical analysis, he creates a delightful perspective on the wonders of whirl. There is even a bonus chapter on how to make simple rotational models, including an entertaining but difficult-to-use drill. Let the good times roll.

The book begins with the lack of macroscopic wheels in biology — an area that Vogel touched on in *Cats' Paws and Catapults* (W. W. Norton, 2000). Notwithstanding whole organisms that tumble and spin, such as the tumbleweed and escaping wheel spiders (*Carparachne aureoflava*), Vogel points out that natural selection has been nearly incapable of producing a freely rotating joint. The only body part that can rotate unimpeded through more than 360 degrees is the bacterial flagellum.

Yet the usefulness and, in many cases, efficiency of rotational movement are such that Vogel proffers many biological examples of how linear motion is translated into rotational, as when contraction of muscle drives rotation of a beater to generate the aptly named *huevos revueltos* (Spanish for scrambled — literally 'revolved' — eggs).

From wheels and carts to bearings and shafts, cranks and drive mechanisms, each chapter is a historically informed circumambulation of an aspect or manifestation of rotation. For instance, which

Why the Wheel Is Round: Muscles, Technology, and How We Make Things Move

STEVEN VOGEL
University of Chicago
Press: 2016.

came first, the cart wheel, or the potter's wheel, used to fashion portable receptacles? (The answer revolves around the bearing.) The many period illustrations

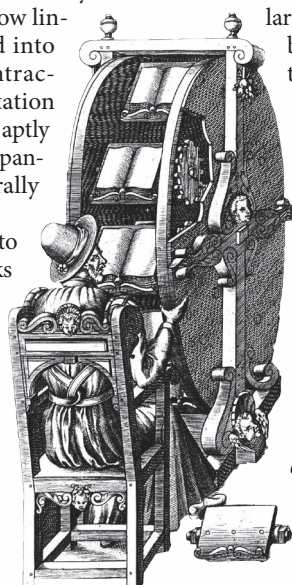
are fascinating: you're compelled to work out the mechanisms of a mule-driven arrastra, a nineteenth-century ore crusher. Some reveal an ingenuity surprising for their era and encourage the reader to appreciate the simplicity with which engineers and designers tackle difficult tasks. The Italian engineer Agostino Ramelli's 1588 picture of an inclined turntable on which a miller would walk in place to turn grinding stones is one such. In other cases, the illustrations challenge your mechanical intuition. Two interlocking elliptical gears from a 1907 image by Gardner Dexter Hiscox look like they will simply jam when rotated. In fact, they turn a constant rotation into an output that is at first faster, then slower, than the input.

From the hidden rotation of a tape measure in its case, through 'true cranks' and treadmills, to how dough creeps up a spinning beater, there are lessons to be learnt. Vogel imparts a cheering numerical understanding, and outlines the possibility of new technologies that leverage under-appreciated concepts. One of these is the zero angular momentum turn — an apparent impossibility demonstrated by a cat righting itself during a fall. No angular momentum is imparted at the outset but, through contortions and twists, the feline reorients along its longitudinal axis.

This marvellous ability has never been exploited technically, but Vogel encourages speculation about use in robotics or micro-electromechanical systems. It might even serve as a metaphor for the elegance and ingenuity that make this book so fun to read. ■

John E. Moalli is a polymer engineer at Stanford University in California. **Adam P. Summers** is a biomechanist at the University of Washington's Friday Harbor Labs.
e-mails: jmoalli@stanford.edu; fishguy@uw.edu

A hypothetical sixteenth-century reading machine.





Ellen Currano (left), Lexi Jamieson Marsh (centre) and photographer Kelsey Vance at a photoshoot.

Q&A Lexi Jamieson Marsh and Ellen Currano

Face to face

Outside the hall containing the posters and exhibits at last month's Geological Society of America meeting in Denver, Colorado, was a surprise. A travelling photography exhibition displayed large, black-and-white portraits of women — wearing beards. To challenge perceptions of who is and is not a scientist, the Bearded Lady Project (www.thebeardedladyproject.com) has photographed more than 75 female Earth scientists; a documentary will be released in early 2017. Filmmaker and project mastermind Lexi Jamieson Marsh and palaeobotanist Ellen Currano of the University of Wyoming in Laramie, who inspired the project, talk about 'invisible women', communities of inclusivity and rocking a moustache.

What prompted this project?

LJM: Ellen and I have been friends for about eight years. We met in the small college town of Oxford, Ohio, where she told me she was a palaeontologist. I was really excited: I had never met one in real life. We were having dinner and Ellen said, "I know how you see me, but I don't necessarily see myself in that light. As a female I'm either very uncomfortable with all eyes being on me for fixing the diversity problems, or I'm ignored, talked over and not paid attention to. There are days I wish I could walk into a room with a beard on my face and just do my work."

EC: It was nothing I had spent a lot of time thinking about. Except that if I were male, my professional life would be easier, because

people would listen to me. There's this celebration of the large grizzled or bearded man going out in the field and facing the elements and being tough and strong, having a large pickaxe and moving giant boulders. And I can't do that. We're not in the documentaries. We're not in *National Geographic*.

LJM: That night, I e-mailed Ellen at 2 in the morning and asked her, what if you did wear a beard? What if we filmed you, and brought in a photographer? And we could nod to the history, that there are no pioneering women palaeontologists in the classic textbooks. We could make up

➔ **NATURE.COM**

For more on science in culture see: nature.com/booksandarts

for a lost legacy and do a tongue-in-cheek response: where are the women if this is the only image we see?

How are you playing on the history of the bearded lady?

LJM: We're walking a fine line, playing with gender identity. I don't expect everyone to understand it. It can be kind of uncomfortable to see a woman with facial hair. But it does nod to the discomfort that comes with women in power positions in science — that they still aren't supposed to be there. The bearded lady is in this ambiguous state of masculine and feminine, which ties nicely in with the many cases of how women in science feel — that they're there, but not really there. With the film, I wanted to challenge what is shown in mass media. It shows women being independent, being physical and scientifically minded.

How did the women choose their facial hair?

LJM: I did most of the facial hair. I have a background in theatre. The scientists choose where they're filmed, what they're photographed wearing, what tools they would like in the picture. The only thing we alter is the beard. We're not dressing them up like men, it's very much who they are and what they do. But if all that changes with the beard, and your mind can't figure out who this person is, that's the goal. Carole Hickman at the University of California, Berkeley, said she would like to participate, but would bring her own moustache. In the 1970s, she worked in the Australian outback and, as a young woman working alone, she was constantly approached by men. She got a moustache, threw it on and got her work done. So that is her moustache.

How do you think you look in the photo?

EC: I think my parents said it best — I look like I should be on a wanted poster. I look tired and run-down, like I've been out in the field for a long time and I'm dirty. And I was. The beard I could do without, but I think I really rock that moustache.

What do you hope people will get out of this project?

EC: The community of inclusion. Making ties between scientists. And the knowledge that you can look however you want and do good science, and people shouldn't be judging you. This project is just one part of getting there.

LJM: I hope it brings awareness that might not be realized in the moment of coming to see the portraits. We want to have something people can think about, and then come to a realization that there is something wrong. The belief system can be questioned. ■

INTERVIEW BY ALEXANDRA WITZ

This interview has been edited for length and clarity.

Correspondence

Two African elephant species, not just one

Your affirmation that the African forest elephant and the African savannah elephant are separate species (*Nature* 537, 7; 2016) is timely. Earlier this month, the 17th Conference of Parties to the Convention on International Trade in Endangered Species (CITES) rejected a proposal to list all African elephants as one species under CITES Appendix I. The US Fish and Wildlife Service is also reviewing a proposal to change the status of both species from threatened to endangered under the US Endangered Species Act (see go.nature.com/2d2ayzb).

Data supporting the separate taxonomic status of African forest elephants (*Loxodonta cyclotis* Matschie) and African savannah elephants (*Loxodonta africana* Blumenbach) have been available for more than a decade. Their evolutionary divergence is comparable in magnitude to that between modern Asian elephants (*Elephas*) and the extinct mammoths (*Mammuthus* spp.). Hybridization between the two African species is rare and highly localized and does not affect the genetic integrity of either species (A. L. Roca *et al.* *Nature Genet.* 37, 96–100; 2004).

In the past decade, African forest elephant populations have fallen by about 60% (T. Breuer *et al.* *Conserv. Biol.* 30, 1019–1026; 2016). Recognition of the forest elephant and the much more numerous savannah elephant as separate species will help to protect their evolutionary diversity.

Colin P. Groves* *Australian National University, Canberra.*
colin.groves@anu.edu.au

*On behalf of 4 correspondents (see go.nature.com/2eye8jf5 for full list).

Include social equity in California Biohub

We have an idea for philanthropists Priscilla Chan and Mark Zuckerberg, who last

month announced their first major investment in basic science: US\$600 million for a Biohub in San Francisco, California.

They aspire to ‘advance human potential and promote equality’ (<https://chanzuckerberg.com>). As members of the Science FARE (Feminist Anti-Racist Equity) collective, we suggest that 5–7% of the Biohub’s health-research budget should be used to design and monitor goals of justice and equality from the outset. Otherwise, social inequalities could limit the project’s potential. Innovative social scientists will need to work with bench scientists, engineers and clinical researchers. Health research should include trained people from all social backgrounds and a variety of disciplines.

The affordability of treatments and access to them is crucial, irrespective of class, gender, race or disabilities. Building equality into Biohub’s founding architecture will allow it to be tackled simultaneously with disease eradication, mitigating the uneven social distribution of health care in San Francisco’s Bay Area and beyond.

Science FARE* *University of California, Berkeley, USA.*
charis@berkeley.edu

*On behalf of 6 correspondents (see go.nature.com/2drsnmf for full list).

Soil clean-up needs cash and clarity

China plans to curb soil pollution by 2020 and to bring environmental risks under control by 2030. In our view, several issues must be addressed for these goals to be realized. Meanwhile, a long-awaited law to prevent soil pollution should be enacted urgently.

The ongoing clean-up requires more funding from local and central government. Polluters should also contribute to remediation costs so that the authorities can decontaminate polluted soil without further liability. Treatment of industrial

sites in inland areas should not be overlooked in favour of megacities in eastern China that have a greater potential for property development.

The administration and supervision of operations needs to be streamlined. Although 36 government departments are involved in soil-pollution control, their respective responsibilities are still not fully defined or coordinated. Standardized regulations must be drawn up to aid communication among stakeholders.

Soil and hydrogeological conditions vary enormously across China, calling for a range of different technologies and skills. International expertise and cooperation could help to address the scientific issues and develop efficient clean-up technologies.

Changsheng Qu, Shui

Wang *Jiangsu Academy of Environmental Sciences, Nanjing, China.*

Peter Engelund Holm *University of Copenhagen, Denmark.*
031202026@163.com

Species loss: learn from health metrics

The inability to quantify which threats matter most across species and ecosystems is a problem for policymaking and resource allocation (see S. L. Maxwell *et al.* *Nature* 536, 143–145; 2016). Biodiversity conservation could learn from public-health metrics and go beyond simply counting the number of recorded threats to quantify the contribution of each one to species loss.

Public-health priorities are set using disability-adjusted life years (DALYs), a measure of healthy years of life lost to a disease as a result of death or sickness. DALYs can be compared among diseases, regions or populations; summed to assess total disease impact; and used to evaluate the effectiveness of interventions (C. J. L. Murray *et al.* *Lancet* 386, 2145–2191; 2015). The absence of these key functions from existing

biodiversity risk assessments limits their usefulness (see, for example, the IUCN Red List).

Although they are not without flaws, DALYs have led to fundamental changes in public health, for example by refocusing efforts on diseases that cause the most harm, such as malaria. They have also prompted reassessment of underlying threats that exacerbate illness, such as malnutrition. And they have highlighted areas in which funding exceeds the share of all DALYs, notably in breast cancer. The availability of an accessible metric, comparable across threats, has also contributed to new funding streams such as the Global Fund to Fight AIDS, Tuberculosis and Malaria.

A comparable metric is urgently needed for more precise analysis of biodiversity threats.

Kathryn J. Fiorella *Cornell University, Ithaca, New York, USA.*
Giovanni Rapacciuolo *Stony Brook University, New York, USA.*
Christopher Trisos *National Socio-Environmental Synthesis Center, Annapolis, Maryland, USA.*
kfiorella@gmail.com

Martian dance of fiction and fact

In marking the H. G. Wells anniversary, you highlight what Carl Sagan dubbed the “dance” between science fiction and science fact (see www.nature.com/scifispecial).

Wells’s *The War of the Worlds* saw the Martian invasion stopped in its tracks by Earth pathogens (S. J. James *Nature* 537, 162–164; 2016). Now, almost 120 years after Wells’s novel was published, the Mars rover Curiosity may have to be diverted because of fears that Earth microbes on the craft could contaminate possible wet areas — a potential habitat for hypothetical Martian life (*Nature* 537, 145–146; 2016). Such symmetry.

Jonathan Cowie *Leicester, UK.*
www.concatenation.org/contact.html

Deborah S. Jin

1968–2016

Pioneer of ultracold quantum physics.

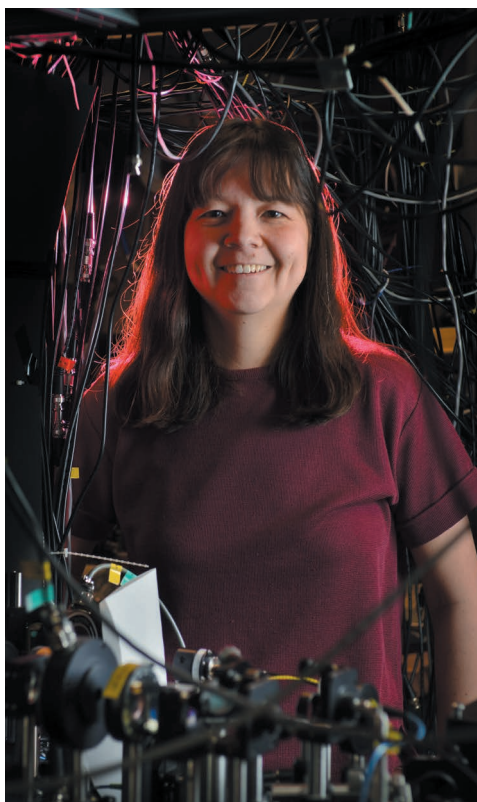
Deborah Jin invented ways to study a state of matter created in the mid-1990s: gases of strongly interacting atoms, cooled to near absolute zero. Her visionary and methodical approach made it possible to use these ultracold gases as model systems to tease out the quantum principles that lead to behaviours in real materials, such as superconductivity.

Jin, who died of cancer on 15 September, aged only 47, was born in 1968 in Stanford, California. She grew up in Indian Harbour Beach, Florida; her father was a professor of physics at Florida Institute of Technology. Her mother and brother also trained as physicists. A studious child, Jin won many mathematics competitions in school.

Jin completed an undergraduate degree in physics at Princeton University in New Jersey in 1990. She earned her PhD in 1995 from the University of Chicago in Illinois under Thomas Rosenbaum, studying unconventional superconductors that were cooled to millikelvin temperatures using liquid helium and similar traditional cryogenic techniques. Jin probed how the exotic superconductivity in these materials reacted to pressure, stress and magnetic fields. During this period, she met her husband, John Bohn, also a physics graduate student at Chicago. In the years that followed, they collaborated on several studies on the collisions of quantum particles.

Next, Jin made the bold decision to change research areas. She moved to Boulder, Colorado, to do postdoctoral research with Eric Cornell at JILA, a joint institute between the National Institute of Standards and Technology (NIST) and the University of Colorado, previously known as the Joint Institute for Laboratory Astrophysics. She worked on materials created with a new set of techniques — quantum gases of atoms cooled with lasers to microkelvin temperatures and suspended in vacuum by magnetic fields.

Jin quickly made key contributions to this new field, including measurements of the heat capacity and the excitation spectra of a Bose–Einstein condensate, a quantum phase of matter that Cornell had created for the first time with a colleague in 1995. Bose–Einstein condensates comprise bosons, particles that have either zero or integer values of spin, that are all in the same spin state. Remarkably, the behaviour of collections of quantum particles, whether in superfluids or neutron stars, is determined



by whether they are made up of bosons or fermions, particles with half-integer spin.

In 1997, Jin accepted a permanent position at JILA and took on one of the greatest challenges in atomic physics at the time — creating a gas of ultracold fermions. Only bosonic gases had been thus cooled when Jin chose to work with the fermionic isotope potassium-40. She realized that trapping two nuclear states of this rare atom was key to producing the first quantum Fermi gas, which her team did in 1999 (B. DeMarco and D. S. Jin *Science* **285**, 1703–1706; 1999). The usefulness of these gases lies in their ability to emulate other models, for example of high-temperature superconductors.

The natural interactions between potassium-40 atoms are too weak to induce strong quantum correlations directly. So Jin and Bohn harnessed collisional resonances to strengthen the interaction between atoms using a magnetic field. They produced the first molecular Fermi condensate in 2003, and the first resonantly interacting Fermi gas in 2004.

Jin's work marked a shift in how other branches of physics viewed and interacted

with experimental atomic physics. The early work on dilute atomic Bose–Einstein condensates could be understood using relatively simple and well-developed theoretical approaches. The achievement of strong interactions in the gas enabled experiments to connect with open and difficult physics questions, some of which still cannot be tackled using the most powerful supercomputers. This led to rich and varied exchanges between research areas, as exemplified by Jin's connections with nuclear and condensed-matter theorists.

In 2008, expanding her work from atoms, Jin partnered with Jun Ye at JILA to create the first quantum gas of diatomic molecules that experience long-range inter-particle interactions. These interactions lead to correlated behaviours in many-body quantum systems. The collaboration also helped to launch the study of chemical reactions in the ultracold quantum realm.

Despite a career cut tragically short, Jin's scientific legacy is broad and deep. Her research made textbook models a scientific reality, including ideal Fermi gases and the crossover between Bose–Einstein condensation and the Bardeen–Cooper–Schrieffer theory of superconductivity. She developed the tools and techniques to tune these states and conduct precise measurements, unveiling new physics along the way.

Debbie inspired a generation of young scientists who founded careers on the research directions she started. She was a warm, dedicated mentor and role model, and a champion of female physicists. She cared deeply about her students, colleagues and friends. Her laser-like focus and intellectual integrity could at times seem blunt, but her ex-students still ask themselves how Debbie would approach a problem.

Debbie loved living in Boulder and exploring the world with her husband and daughter, Jaclyn, who was often to be seen in the background at scientific meetings. Her bright smile will be missed by the many people whose lives she touched. ■

Brian DeMarco is a professor of physics at the University of Illinois Urbana–Champaign, Urbana, Illinois, USA. He was Debbie's first graduate student and worked with her on creating the first Fermi gas.

John Bohn, Debbie's husband, and **Eric Cornell**, a collaborator, are fellows at JILA. e-mail: bdemarco@illinois.edu

GLENN ASAKAWA/UNIV. COLORADO

GENOMICS

A matched set of frog sequences

A whole-genome duplication that occurred around 34 million years ago in the frog *Xenopus laevis* made generating a genome sequence for this valuable model organism a challenge. This obstacle has finally been overcome. [SEE ARTICLE P.336](#)

SHAWN BURGESS

Ask a developmental biologist to name the most valuable animal models for their field and they will probably put the African clawed frog, *Xenopus laevis*, at or near the top of their list. Ask any geneticist the same question and this species is unlikely to even make the top ten. One reason for this disparity is that *X. laevis* has undergone a whole-genome duplication, which makes genome assembly — an essential tool of modern genetics — extremely difficult. But on page 336, Session *et al.*¹ report the successful sequencing of the *X. laevis* genome. The authors took advantage of ever-improving technologies and the hard work of a large, international consortium to complete this challenging project.

During the genome-assembly process for a diploid organism (one, like humans, that has two sets of chromosomes), a single reference chromosome sequence is generated to correspond to each chromosome pair. *X. laevis*, by contrast, is tetraploid — it has four sets of chromosomes, and so a reference sequence will contain two copies of most genes, instead of one. This leads to problems when using the typical shotgun approach to genome assembly, in which hundreds of millions of random short sequence reads are taken and assembled by computer into logical, continuous sequences. With a duplicated genome, it can be difficult to tell which of the two gene copies a short sequence comes from. If the sequences of the copies are too similar, the computer's assembly algorithms 'collapse' the duplicated sequence into a single copy, confounding attempts to make correct, end-to-end assemblies across all chromosomes.

Two laborious approaches that enable distinctions between duplicated chromosomes made Session and colleagues' effort successful. In the first, the authors isolated DNA from a frog and inserted long stretches into DNA constructs called bacterial artificial chromosomes (BACs). They then systematically identified 798 BACs that contained large fragments (100 kilobases or more) of DNA encoding one copy of a duplicated gene, and that could be paired with another BAC containing the other copy.

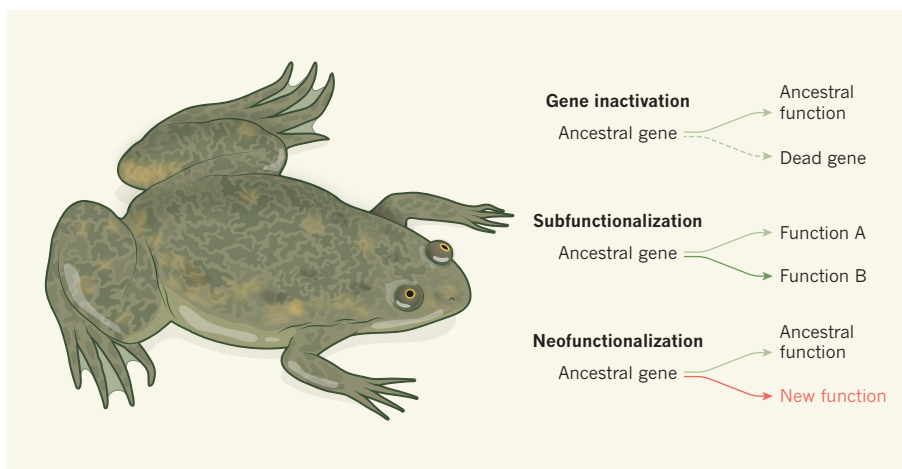


Figure 1 | Pathways to rediploidization. After an organism undergoes a whole-genome duplication (WGD), as has occurred in the frog *Xenopus laevis*, its genome will slowly return to the normal diploid state (in which it inherits just one set of chromosomes from each parent) through three mechanisms. First, the most common outcome is for one copy of an ancestral gene to pick up enough mutations that it becomes inactivated and 'dies'. Second, if a gene has multiple functions, these roles can be split between the two copies (subfunctionalization), and both copies will be maintained. Third, a less common outcome is neofunctionalization, in which one gene duplicate evolves a new function. By sequencing the genome of *X. laevis*, which has undergone a WGD, Session *et al.*¹ investigate these processes. (Graphic adapted from an original by Darryl Leja.)

The researchers used these paired BACs to make pairs of DNA 'probes', which bind to the two DNA sequences and are each labelled with a different fluorescent molecule, and then simultaneously hybridized the two probes to intact chromosomes. This process allowed them to assign the two duplicated genes to the correct chromosomes on the basis of which colour probe bound specifically to which chromosome, effectively re-separating collapsed sequences. The technique improved genome assembly by enabling stretches of assembled sequence to be strung together into larger, chromosome-assigned chunks.

The second technique was tethered conformation capture, in which regions of tightly packaged DNA are cross-linked together and the joined DNA pieces are subsequently sequenced as a pair. Most cross-linked sequence pairs come from the same chromosome, but can be up to hundreds of kilobases apart. As such, sequences can be linked to others from the same chromosome, creating larger continuous sequences. Together, these two techniques enabled the separation

of duplicated sequences into distinct chromosomes, resulting in a high-quality genome sequence.

We could stop here and Session and colleagues' work would already be of major interest. But species that have undergone a whole-genome duplication (WGD) also provide an opportunity to watch evolution happen within a species, instead of piecing together evolutionary paths by comparing distinct species. After a WGD, duplicated genes that share the same function can undergo several types of change over time: inactivating mutations can arise in one copy (the mutated copy 'dies'); the original functions can be split between the two copies; or one copy can develop a new function while the other gene retains its ancestral role (Fig. 1). Given enough time, the organism will return to a diploid state, in which all remaining genes have unique and evolutionarily important functions. This process of rediploidization has happened several times during vertebrate evolution².

Xenopus is the third vertebrate with a duplicated genome to be sequenced in the past three

years, joining the common carp (*Cyprinus carpio*)³ and the Atlantic salmon (*Salmo salar*)⁴. Of the three, the WGD in carp occurred most recently, only 8 million years ago. The Atlantic salmon genome was duplicated 80 million years ago, and the *X. laevis* genome provides us with an intermediate, at approximately 34 million years.

Session and colleagues made some interesting observations when looking at within-species evolution in *X. laevis*. They found that protein-coding genes were retained at a higher rate than expected, suggesting that maintaining balanced expression levels is necessary for more genes than previously thought. Conversely, conserved non-coding elements (CNEs) — the regions of the genome most likely to be sequences such as enhancers or promoters that regulate gene expression — were retained at a significantly lower rate. This fits with the idea that regulatory elements have more freedom to change in a duplicated genome, accelerating evolution.

Another interesting phenomenon was that one paired set of chromosomes (dubbed S) was almost four times more likely to have a gene die or be deleted than the other (X). It is unclear why this would occur — perhaps certain aspects of the physiology of the new frog species that emerged from the WGD were more compatible with the ancestor that contributed the X chromosomes, thus favouring retention of this set. Session *et al.* also noticed that certain categories of gene were more likely to be specifically retained in two copies — in particular, those encoding DNA-binding proteins and proteins of developmentally regulated signalling pathways. One reason given by the authors for this is that transcription factors and signalling molecules that often rely on gradients of expression for their effect on development might be more sensitive to alterations in copy number than most other proteins.

Finally, Session and colleagues showed that many pairs of duplicated *X. laevis* genes have divergent spatio-temporal expression. These alterations in gene expression are a good opportunity for scientists to connect gene expression to the molecular evolution of duplicated CNEs. In other words, alterations in enhancer sequences can be correlated with alterations in gene expression, and the causative proof of these changes is particularly compelling because the changes can be measured against another copy of the same gene within the same species.

The genome sequence for the African clawed frog gives us much to celebrate. Developmental biologists now have at their disposal the detailed genomic information so essential to modern biology. Genome biologists have proof that even large, complex genome duplications can ultimately be resolved into high-quality assemblies. And evolutionary

biologists have another powerful tool with which to examine the birth and death of genes and their regulatory elements during evolution. *Xenopus* has made a huge leap forward as a model organism — scientists will surely follow. ■

Shawn Burgess is in the Translational and Functional Genomics Branch, National

Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

e-mail: burgess@mail.nih.gov

1. Session, A. M. *et al.* *Nature* **538**, 336–343 (2016).
2. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
3. Xu, P. *et al.* *Nature Genet.* **46**, 1212–1219 (2014).
4. Lien, S. *et al.* *Nature* **533**, 200–205 (2016).

ASTROPHYSICS

Unexpected X-ray flares

Two sources of highly energetic flares have been discovered in archival X-ray data of 70 nearby galaxies. These flares have an undetermined origin and might represent previously unknown astrophysical phenomena. [SEE LETTER P.356](#)

SERGIO CAMPANA

Since our beginning, humanity has observed and chronicled stars, comets and supernova explosions. Nowadays, we have instruments that can cover a wide range of the electromagnetic spectrum, from low-energy radio waves to high-energy X-rays and γ -rays. The high-energy sky is the theatre for many time-varying and energetic phenomena, often involving compact objects such as neutron stars or black holes. On page 356, Irwin *et al.*¹ report what looks like a previously

undescribed kind of bright X-ray flaring activity in neighbouring galaxies.

In 2005, an unexplained source of X-ray flares was discovered² in proximity to the galaxy NGC 4697. Motivated by this discovery, Irwin and colleagues analysed archival X-ray data from the Chandra and XMM-Newton space observatories and found two further X-ray sources that exhibit particularly bright flares. The authors' sources are located in the outskirts of nearby galaxies: one near the Virgo galaxy NGC 4636 and the other near the Centaurus A galaxy (NGC 5128; Fig. 1). The

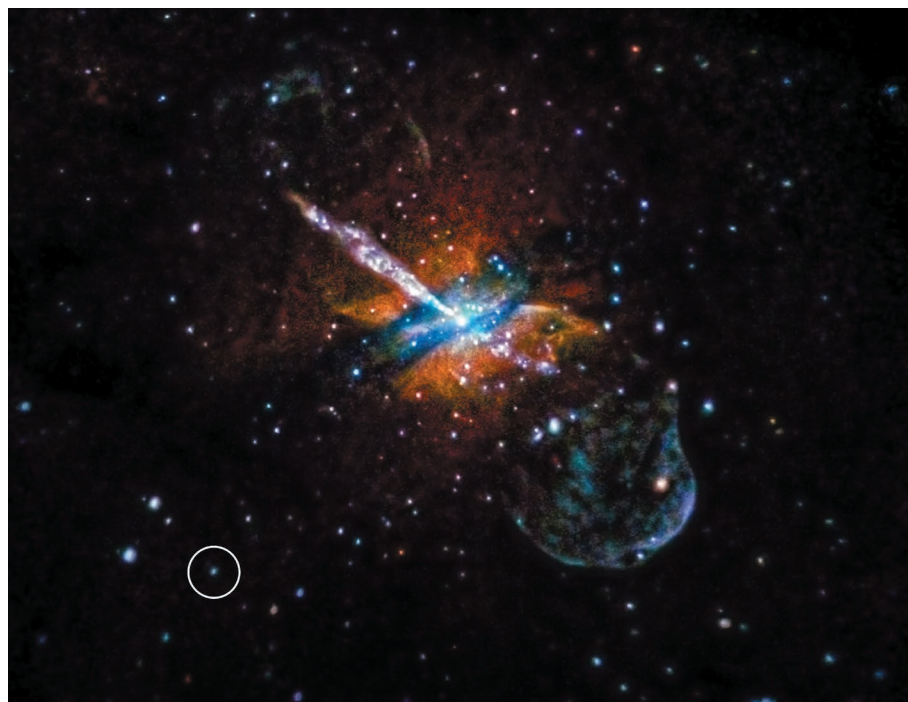


Figure 1 | The environment of Centaurus A. Irwin *et al.*¹ observe a source of ultraluminous X-ray flares (indicated by the white circle) near the elliptical galaxy Centaurus A (NGC 5128; centre).

NASA/CXC/UNIV. BIRMINGHAM/M. BURKE ET AL.

archival data for the latter source revealed five different flaring events. The peak luminosity of the flares (up to 10^{34} watts) is intriguing, because it is even greater than the luminosity that can be attained by a neutron star under normal circumstances. All the flares had short rise times — they reached their peak luminosities in less than a minute — and lasted for about an hour.

What is the origin of these flares? In general, such events can be identified by their duration and whether or not they recur. Non-recursive events that have durations similar to the observed flares include long-duration γ -ray bursts³ and explosions that signal the deaths of massive stars (energetic supernovae). However, these explanations require a population of young stars, whereas the sources are located in the outskirts of elliptical galaxies, which would indicate an old stellar population.

Repeated bursts of high-energy radiation are observed from the flaring of stars that are less luminous than our Sun. Meanwhile, neutron stars are generous sources of different kinds of burst. Highly magnetic neutron stars called magnetars in our Galaxy can give rise to bright flares that last for less than a second⁴, but can still influence Earth's ionosphere. By contrast, neutron stars that accrete matter from a companion star can produce 'type-I' X-ray bursts, resulting from runaway thermonuclear combustion of matter accumulated on the neutron star's surface. These bursts repeat, but they cannot explain the observed signals because they last for only a few minutes and are less luminous by a factor of about 100 (ref. 5). Finally, neutron stars can also exhibit super-bursts, which are longer versions of type-I X-ray bursts (lasting for a few hours⁶), but these are also under-luminous compared with the flares discovered here. Therefore, even within this zoo of different transient phenomena, Irwin and colleagues' flaring sources remain unmatched.

Do we have any other information? Both sources are located in old stellar populations, which probably rules out the presence of young neutron stars such as magnetars. In addition, the sources reside in what seem to be globular clusters or ultracompact dwarf galaxies. The optical spectrum from one of the sources shows that it is at the same distance from Earth as its host galaxy, which confirms that we are not looking at a star in the Milky Way that is being projected onto a more distant galaxy. Both sources also have a high X-ray luminosity before and after their flares that is larger than can be achieved by a neutron star (even if exceptions exist for young neutron stars⁷). All of these observations seem to suggest the presence of a black hole.

If the black-hole interpretation is correct, there are at least two viable explanations. One possibility, which is suggested by the authors, is that there exists an intermediate-mass black hole (100–1,000 times the mass of the Sun)

at the centre of each source that, for some unknown reason, emits flares that last for about an hour. Alternatively, one could envisage a lower-mass black hole whose X-ray emissions are beamed directly towards Earth. A binary system that has a highly eccentric orbit⁸ might explain the repeated flares detected from the source near NGC 5128). The flares from such a binary system would be strictly periodic, because sporadic surges of accretion would occur at the point of closest approach.

The mystery of Irwin and collaborators' flares might be solved, as often happens in astrophysics, with more observations. In particular, one could investigate the frequency at which the flares occur. An important aspect of the authors' discovery is that both sources were identified after a careful search through archival data, which provides further testimony for the legacy value of such archives. Now that

we know these strange objects are out there, they will remain on the watch list and more examples will be searched for. ■

Sergio Campana is at the *Istituto Nazionale di Astrofisica (INAF), Osservatorio Astronomico di Brera, 23807 Merate, Italy.*
e-mail: sergio.campana@brera.inaf.it

1. Irwin, J. A. *et al.* *Nature* **538**, 356–358 (2016).
2. Sivakoff, G. R., Sarazin, C. L. & Jordán, A. *Astrophys. J.* **624**, L17–L20 (2005).
3. Gehrels, N., Ramirez-Ruiz, E. & Fox, D. B. *Annu. Rev. Astron. Astrophys.* **47**, 567–617 (2009).
4. Palmer, D. M. *et al.* *Nature* **434**, 1107–1109 (2005).
5. Strohmayer, T. & Bildsten, L. in *Compact Stellar X-ray Sources* (eds Lewin, W. & van der Klis, M.) 113–156 (Cambridge Univ. Press, 2006).
6. Cornelisse, R., Heise, J., Kuulkers, E., Verbunt, F. & in't Zand, J. J. M. *Astron. Astrophys.* **357**, L21–L24 (2000).
7. Bachetti, M. *et al.* *Nature* **514**, 202–204 (2014).
8. Maccarone, T. J. *Mon. Not. R. Astron. Soc.* **364**, 971–976 (2005).

CANCER

A shocking protein complex

Heat-shock proteins have been found to form part of a large protein complex, called the epichaperome, that improves the survival of some cancer cells. This complex might offer a new target for cancer treatment. SEE LETTER P.397

KAI BARTKOWIAK & KLAUS PANTEL

Tumour cells are subject to various forms of stress, such as oxygen or nutrient shortages, when tumour blood-vessel formation cannot keep pace with tumour growth — and the cells therefore need to have effective stress-survival strategies. Heat-shock proteins (HSPs) are often active in cells exposed to stressful conditions. On page 397, Rodina *et al.*¹ have investigated the role of HSPs in human cancer, and find that the proteins exist in a large complex in some cancer cells.

Maintenance of the correct 3D structure of a protein is essential for its function, and cells have mechanisms dedicated to protein quality control. Proteins with structural defects are either refolded into the correct conformation or, in the case of severe structural abnormality, targeted for degradation. Protein quality control is regulated by chaperone proteins, which are often involved in facilitating folding. Chaperones include HSPs, which are active in human cells both under normal conditions and in stressful conditions such as inflammation and a state of reduced oxygen known as hypoxia.

Rodina and colleagues investigated HSPs in human cancer cells using a biochemical technique that separates proteins in samples

of cellular proteins. The authors found that, in samples of what they call 'type 2' cancer cells, the heat-shock protein HSP90 separated as expected on the basis of its structure. However, in other samples tested, HSP90 had an unexpected separation pattern that could be explained by its presence in a large complex with other HSPs. The authors use the term 'type 1' cancer cells to describe cells containing this large protein complex (Fig. 1).

The researchers found that, in type 1 cancer cells, HSP90 associated with dozens of other proteins, including scaffolding and adaptor proteins, which are also involved in regulating protein folding. The authors called this structure the epichaperome. By contrast, in type 2 cancer cells and non-cancer cells, HSP90 was associated with only a small set of proteins, and most of the HSPs existed as solitary proteins or were assembled into small complexes.

HSP90 requires the essential nucleotide ATP to function. The authors therefore targeted the epichaperome using the molecule PU-H71, which binds to HSP90's ATP-binding pocket and inhibits the protein's function. The inhibitor bound more tightly when HSP90 was in the epichaperome, and killed more type 1 cancer cells than type 2 or non-cancer cells. This difference was not attributable solely to chemical inactivation of HSP90, because the authors

found that selective genetic downregulation of other HSPs resulted in increased death of type 1 cells compared with type 2 cells, suggesting that type 1 cell survival depends on an intact epichaperome.

Using protein-complex analysis techniques and PU-H71 treatment, Rodina and colleagues investigated the epichaperome in different types of cancer cell. The epichaperome was detected in 60–70% of cell lines from breast, pancreatic, lung, leukaemia and other cancers, indicating the potential clinical relevance of this protein complex as a therapeutic target. Strikingly, the epichaperome was not restricted to a particular cancer subset characterized by specific gene or protein expression. Type 1 cells also had high levels of the cancer-associated MYC protein. When MYC was downregulated, the epichaperome disappeared, whereas overexpressing MYC in type 2 cells induced epichaperome formation. Thus, MYC is an important regulator of epichaperome assembly.

Treatments for cancer require a target that is present in most cancer cells but absent in normal cells. Rodina and colleagues' finding that type 1 cancer cells contain HSP90 in the epichaperome, but that normal cells have HSP90 mainly in a non-complexed form, suggests that the epichaperome might be a clinical target.

Despite the encouraging results, the authors found considerable differences in dependence on epichaperome formation, both between tumour types and between tumour cells of the same type (for example, breast cancer), which might allow the emergence of treatment-resistant variants. Moreover, in many cancers, such as breast cancer, the metastatic migration of cancer cells to other locations in the body is the main cause of death, and metastatic cells may differ from the initial primary tumours. It remains to be determined to what extent metastatic cells depend on epichaperome formation, and how such cells might respond to drugs that target the complex.

However, obtaining metastatic tissue from a patient by needle biopsy is a problem because of the procedure's invasive nature, and some lesions (for example, in the lung, bone and brain) are difficult to access. 'Liquid biopsy' of circulating tumour cells in the peripheral blood might be an alternative feasible strategy for assessing the epichaperome in metastatic cells. Analysis of circulating tumour cells or tumour cells that have migrated to the bone marrow in people in the early stages of cancer might also provide information on potential precursors of metastases^{2,3}.

Rodina and colleagues' findings could have broader implications. Many cellular programs consist of signalling pathways that rely on a large number of proteins, and these might form stable complexes in a similar way to how the epichaperome forms. For example, the chaperone machinery in a cellular organelle

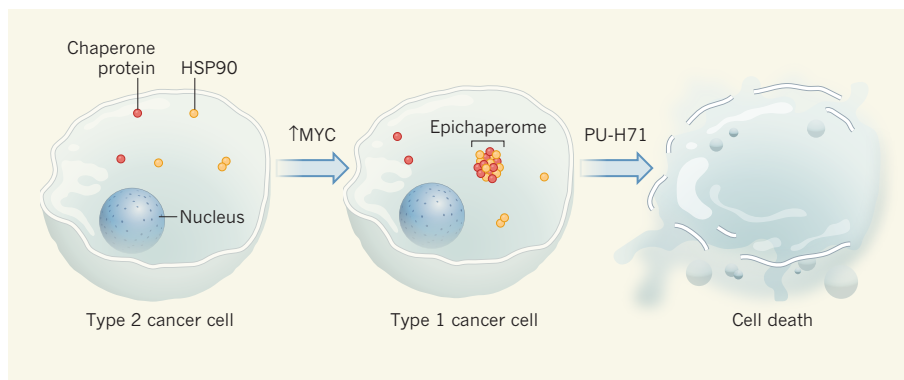


Figure 1 | The epichaperome in cancer cells. Rodina *et al.*¹ have found that, in certain cancer cells (which the authors call type 2 cells), heat-shock proteins such as HSP90, and other proteins associated with the response to cellular stress such as chaperones, are present as solitary proteins or in small complexes. However, other cancer cells (referred to as type 1 cells) show the formation of a large protein complex called the epichaperome, which contains HSP90 and other proteins, including chaperones. Increasing the levels of MYC protein in type 2 cells induced the formation of the epichaperome. When type 1 cells were treated with the HSP90 inhibitor PU-H71 *in vitro*, the epichaperome disintegrated and the cells died. The epichaperome might thus represent a target for treating cancer.

called the endoplasmic reticulum can form large protein complexes⁴, and the activity of endoplasmic reticulum chaperones under conditions of cellular stress strongly affects cancer-cell survival⁵. Moreover, cancer-associated ErbB receptors might assemble into higher-order receptor complexes⁶. Thus, the discovery of other large, cancer-specific protein complexes might open further avenues of investigation for understanding cancer biology, with potential implications for the design of new strategies for cancer therapy. ■

Kai Bartkowiak and Klaus Pantel are in

the Department of Tumor Biology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. e-mail: pantel@uke.de

1. Rodina, A. *et al.* *Nature* **538**, 397–401 (2016).
2. Alix-Panabières, C. & Pantel, K. *Nature Rev. Cancer* **14**, 623–631 (2014).
3. Bartkowiak, K. *et al.* *Cancer Res.* **75**, 5367–5377 (2015).
4. Meunier, L., Usherwood, Y.-K., Chung, K. T. & Hendershot, L. M. *Mol. Biol. Cell* **13**, 4456–4469 (2002).
5. Lee, A. S. *Nature Rev. Cancer* **14**, 263–276 (2014).
6. Sweeney, C. & Carraway, K. L. III *Oncogene* **19**, 5568–5573 (2000).

This article was published online on 5 October 2016.

DRUG DISCOVERY

Chemical diversity targets malaria

A molecule selected from a library of compounds that have structures similar to natural products targets several stages of the malarial parasite's life cycle, offering single-dose treatment of the disease in mouse models. [SEE ARTICLE P.344](#)

DAVID A. FIDOCK

Eliminating malaria would save more than 400,000 lives annually, mainly those of young children in sub-Saharan Africa, and prevent the approximately 200 million cases of the disease that arise each year¹. Achieving this will require medicines that can eliminate all three stages of *Plasmodium* parasite infection in humans. On page 344, Kato *et al.*² report compounds, known as bicyclic azetidines, that display this multistage activity.

Kato and colleagues' work comes at a pivotal time. Malaria treatments rely on artemisinin-based combination therapies (ACTs), which combine an artemisinin-based compound with a second antimalarial drug. The rapid efficacy and global implementation of ACTs, coupled with increased mosquito-vector control efforts, have halved malaria death rates in the past 15 years¹. However, artemisinin resistance has emerged and is now widespread in southeast Asia³. This places the partner drugs under increased evolutionary selection pressure for the development of resistance. In

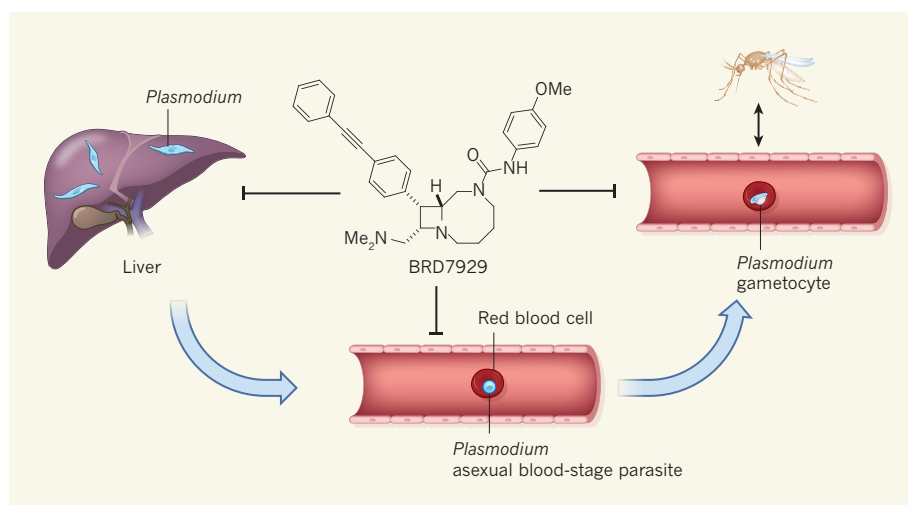


Figure 1 | A compound that targets all parasite stages of malarial infection. Malaria is caused by infection with *Plasmodium* parasites that progress through three main stages in the host. The first stage is an asymptomatic liver infection. Then parasites infect red blood cells at the disease-causing asexual blood stage. Parasites can also form gametocytes inside red blood cells, which can be taken up from the bloodstream by mosquitoes to transmit the disease onwards. Current artemisinin-based malarial treatments target asexual blood stages and immature gametocytes. By screening a chemical library for compounds active against different stages of the *Plasmodium* parasite's life cycle, Kato *et al.*² identified the compound BRD7929, which, when tested against the human parasite *Plasmodium falciparum* or the rodent species *Plasmodium berghei* in mouse models, provided a single-dose cure and was able to prevent malaria and block onward transmission of the parasite. Me, methyl group.

Cambodia, the ACT of dihydroartemisinin and piperazine is failing rapidly because of resistance to both compounds⁴. Historical precedents suggest that resistance to first-line ACT agents might take hold in Africa next. There is thus an urgent need to develop more medicines.

The first stage of *Plasmodium* infection in humans is an asymptomatic infection of the liver; the second stage occurs when asexual parasites cause disease by infecting red blood cells; and at the third stage, parasites form sexual-stage gametocytes in red blood cells that, once mature, can be transmitted to *Anopheles* mosquitoes (Fig. 1). To identify potential multistage inhibitors, Kato *et al.* first carried out a high-throughput screen using *in vitro* cultures of asexual blood-stage parasites of *Plasmodium falciparum*, the most lethal of the human malarial parasites.

The authors tested a library of approximately 100,000 compounds to search for parasite growth inhibitors. These compounds were produced using a technique known as diversity-oriented synthesis, in which chemical structures are built and coupled to generate molecules in a process inspired by the diversity and structural complexity of naturally occurring compounds⁵. Kato and colleagues then tested their inhibitory compounds against asexual blood-stage parasites from a panel of parasite strains resistant to known antimalarial agents, and extended the assays to liver and gametocyte-stage parasites.

The authors' screens identified several series of 'hits' that acted on known malarial-drug-target proteins, such as *P. falciparum*

ATP4, PI4K and DHODH (ref. 6), as well as a wealth of other hits (including bicyclic azetidines) that have potentially new modes of action. The hits are documented at the Malaria Therapeutics Response Portal website (<http://portals.broadinstitute.org/mtrp>), which is a valuable resource for future antimalarial-drug discovery and development efforts.

By selecting for drug resistance in cultured parasites and applying whole-genome DNA-sequence analysis to identify genetic changes, Kato *et al.* obtained evidence that the bicyclic azetidines target an enzyme termed cytosolic *P. falciparum* phenylalanyl-tRNA synthetase (*Pf* PheRS). This finding was confirmed in biochemical assays demonstrating chemical inhibition of the enzyme. PheRS acts on transfer-RNA molecules, enabling them to deliver the amino acid phenylalanine to nascent proteins during the vital cellular process of messenger-RNA translation and protein synthesis. tRNA synthetase enzymes have emerged in recent years as a promising class of antimalarial target⁷.

Kato *et al.* tested their lead bicyclic azetidine compound BRD7929 in mouse models of malarial infection using either *P. falciparum* or the rodent parasite, *Plasmodium berghei*. The authors found that a single, low dose of BRD7929 was sufficient to eliminate infections at the liver or asexual blood stages, affording complete cure. Transmission-blocking activity at the gametocyte stage was also observed at drug concentrations that achieved single-dose cures of asexual blood-stage infections. If similar efficacy could be achieved in treating human infections, it would change the

landscape of disease treatment and control by providing a powerful tool for malaria elimination⁶.

The inspired decision by Kato and colleagues to screen natural-product-like compounds amenable to synthesis was one key to the study's success. The other notable factor was the remarkable coordination between the various collaborating research laboratories. This enabled testing of *Pf* PheRS inhibitors throughout the parasitic life cycle, as well as chemical optimization and pharmacological evaluation of the compounds that was achieved by assessing the relationship between their structures and activities. Of course, there is no guarantee that bicyclic azetidines will ultimately yield a licensed medicine with the desired single-dose cure, prevention and transmission-blocking properties.

BRD7929 displayed good oral bioavailability (the amount of drug that reaches the bloodstream after oral ingestion) and other promising pharmacological properties, including a long half-life — approximately 32 hours in mice. Yet it remains possible that this compound might encounter setbacks during further testing, such as toxicological issues or problems in drug selectivity for the parasite enzyme over its human counterpart. The establishment by Kato *et al.* of a functional screen that uses *Pf* PheRS provides opportunities to identify alternative chemical scaffolds to bicyclic azetidines, if necessary.

Given the constant concern of antimalarial drug resistance⁸, Kato and colleagues tested their compound for resistance development. They found that bicyclic azetidine treatment selects for resistance with a frequency of greater than 1 per 10⁹ parasites, which compares favourably with other preclinical drug candidates^{9,10}. To place this in context, an infected, symptomatic individual can harbour up to 10¹² parasites⁶. Therefore, resistance might emerge in settings in which the disease is endemic, although its ability to become established and spread is tempered by many factors, including host immunity and parasite fitness. Resistance concerns can be mitigated by combining PheRS inhibitors with pharmacologically matched inhibitors that have a different mode of action.

Future studies will also be needed to evaluate whether PheRS inhibitors can effectively eliminate another human malaria parasite, *Plasmodium vivax*, which has a dormant stage of liver infection. These dormant parasites can reinstate growth and cause relapse months or even years after the primary infection. The need for a drug to target dormant *P. vivax* liver infections is particularly acute because primaquine, the only licensed drug treatment for this form of the disease, can be highly toxic to people with certain types of deficiency for the enzyme glucose-6-phosphate dehydrogenase — a common genetic trait⁶.

Single-dose malaria cures, combined with

safe and effective prevention and transmission-blocking measures, would be a tremendous boon to afflicted populations, whose socio-economic challenges are exacerbated by this debilitating disease. The world would be a healthier community of nations if this goal were to be achieved. The current study demonstrates the power of coordinating the multifaceted research activities required to achieve such a goal. ■

David A. Fidock is in the Departments

of Microbiology and Immunology and of Medicine, Division of Infectious Diseases, Columbia University Medical Center, New York, New York 10032, USA.
e-mail: df2260@cumc.columbia.edu

1. WHO. *World Malaria Report 2015* www.who.int/malaria/publications/world-malaria-report-2015/report/en (2015).
2. Kato, N. et al. *Nature* **538**, 344–349 (2016).
3. Ashley, E. A. et al. *N. Engl. J. Med.* **371**, 411–423 (2014).
4. Leang, R. et al. *Antimicrob. Ag. Chemother.* **59**, 4719–4726 (2015).

5. Lowe, J. T. et al. *J. Org. Chem.* **77**, 7187–7211 (2012).
6. Wells, T. N. C., Hooft van Huijsduijnen, R. & Van Voorhis, W. C. *Nature Rev. Drug Discov.* **14**, 424–442 (2015).
7. Pham, J. S. et al. *Int. J. Parasitol. Drugs Drug Res.* **4**, 1–13 (2014).
8. Corey, V. C. et al. *Nature Commun.* **7**, 11901 (2016).
9. Baragaña, B. et al. *Nature* **522**, 315–320 (2015).
10. Phillips, M. A. et al. *Sci. Transl. Med.* **7**, 296ra111 (2015).

This article was published online on 12 October 2016.

OPTICAL PHYSICS

Speedy electrons exposed in a flash

A link has been established between high-frequency light emissions and electron oscillations induced in an insulator by a laser. This is a key step in efforts to make electronic devices that work faster than is currently possible. SEE LETTER P.359

MICHAEL CHINI

The speed at which electronic devices can function is determined by the frequency at which alternating electric currents can be driven in the device. One approach that might surpass the present frequency limit is to drive currents using the nonlinear response of electrons to the oscillating electric field of light. On page 359, Garg et al.¹ report key advances in efforts to realize this goal: the use of intense laser pulses to induce currents controllably at frequencies more than 100 times higher than the present limit, and an approach that allows the associated electron oscillations to be measured.

Optics and electronics have always been closely intertwined — light is, after all, an electromagnetic wave. Our ability to measure and control light fields has advanced tremendously in recent decades, and laser-based telecommunication has been commonplace since the late 1990s. By contrast, there are still gaps in our ability to control alternating electronic currents, which remain limited to frequencies of about 100 gigahertz (1 GHz is 10^9 Hz). Finding a way to use the response of electrons to the electric-field oscillations of a strong light field (several hundred terahertz; 1 THz is 10^{12} Hz) to drive alternating electric currents at even higher frequencies has been an elusive goal.

High-speed circuits rely on the fast conversion of a (typically semiconducting) material from an insulating to a conducting state. The conversion is associated with electrons jumping between energy bands in the material — that is, from the valence to the conduction band. Once in the conduction band,

electrons can be steered by light or by a voltage, resulting in an electric current.

The timescale for the conversion is set by quantum mechanics: insulators and wide-band-gap semiconductors can undergo switching between energy bands at high speeds because of the large energy separation between the valence and conduction bands. It is perhaps unsurprising, then, that the excitation of charge carriers in a bulk insulator (such as the silica nanofilms used by Garg et al.) might

enable the generation of high-frequency currents. In fact, evidence for switching speeds close to 1 PHz (10^{15} Hz) was previously reported by researchers from the same institution^{2,3}. The next challenge was to measure the oscillatory motion of the electrons, and thereby characterize the frequency of the resulting currents.

Garg et al. addressed this challenge starting from the realization that accelerating electrons can emit light known as high-order harmonics⁴, which directly reflect the motion of the oscillating electrons. To link this motion to high-frequency currents, it is first necessary to prove that high-order harmonics are generated only from electron motion within the conduction band, and not from electrons falling from the conduction to the valence band. The authors did this by measuring the relative timing of the different frequencies of light emitted from a silica nanofilm, using a device known as an attosecond streak camera.

The researchers observed that the generated light is emitted in a single burst lasting less than 500 attoseconds (1 as is 10^{-18} s), and

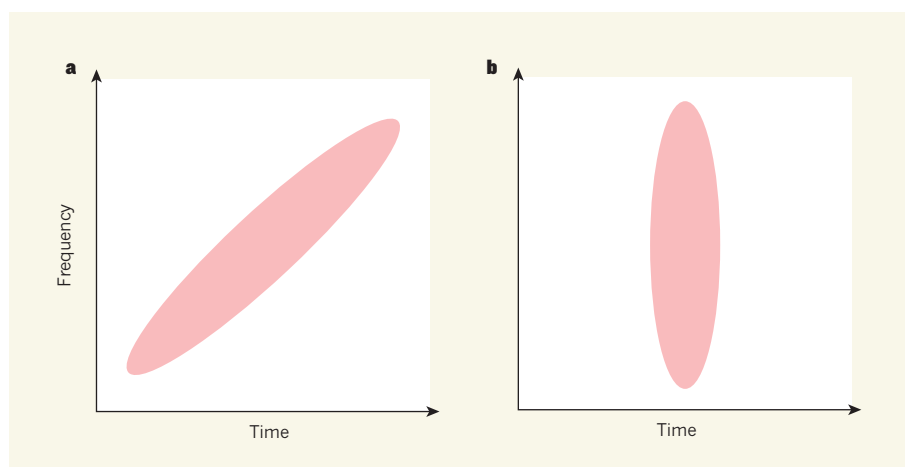


Figure 1 | Differences between high-order harmonic emissions from solids. When solids are irradiated with intense light pulses, electrons jump between the material's energy bands, from the valence band to the conduction band, and can go on to emit light known as high-order harmonics. **a**, If the electrons emit light by dropping back to the valence band, the time taken between the excitation and light emission is longer for high-frequency light than for low-frequency light. **b**, Alternatively, electrons can generate light while remaining in the conduction band. In this case, emissions of different frequencies are all produced at the same time. Garg et al.¹ report that the second case is true for high-order harmonics produced when a silica nanofilm is irradiated using an optical laser — a finding that opens up opportunities for measuring electrons oscillating at multi-petahertz frequencies (1 PHz is 10^{15} Hz).



50 Years Ago

The question of how cells estimate their location within the body is closely related to that of why cells of a developing organism become differentiated... We are now investigating the mechanism involved... using abdominal segments of the pupa *Galleria mellonella*. A previous investigation showed that the scale patterns in *Galleria*¹... are oriented by a concentration gradient of an unknown diffusible substance³. The substance seems to be produced at one margin of the segment and destroyed at the other⁴... To move the concentration gradient to some other part of the segment, pieces of skin were rotated in the larvae by 180°... The concentration gradient, the existence of which is confirmed by these results, obviously has two functions: (1) to orientate the scales by its direction, (2) to supply the cells... with necessary information about their distance from segment margins and to induce the corresponding cuticular structures. **From Nature 22 October 1966**

100 Years Ago

The Psychology of Relaxation. By Prof. G. T. Patrick... In the author's view... forms of human behaviour are, at bottom, illustrative of a single principle. The activities and relations of civilised life imply the upbuilding and functioning of extremely complex mental mechanisms full of tensions, restraints, and inhibitions. To maintain these always in operation is an impossible task. From time to time, therefore, the complexes break up, and man falls back with relief into conduct expressive of simpler mental structures organised and consolidated in the far distant days of the race's childhood: he plays, he laughs, he swears, he fights. **From Nature 19 October 1916**

that there is almost no delay between emissions produced at different frequencies. These observations clearly agree with models in which electron motion occurs in a single band (Fig. 1). Electrons moving between bands would instead result in a 'chirped' emission, in which high-frequency light is emitted later than low-frequency light.

The findings provide links between a study² that separately demonstrated laser-induced, high-speed switching of an insulator between conducting states, and an investigation⁵ that reported high-frequency light emission from laser-irradiated insulators. In other words, the new results show that the phenomena reported in those previous studies originate from the response of electrons in the conduction band to laser pulses. The results also open the way to the use of high-order harmonics as a tool for electronic metrology. Furthermore, Garg *et al.* report that the observed response of electrons to the strong light field is extremely nonlinear: the emitted light extends into the extreme ultraviolet region of the spectrum, which is more than ten times the frequency of the driving light field, and corresponds to energies nearly three times that of the band gap of silica. This extends the range of frequencies at which electronic measurements can be made to well beyond the frequency of the laser light, into the multi-petahertz regime.

Harnessing the potential of optically induced currents and multi-petahertz electronic metrology will be challenging. The observed light emissions, and the currents from which they are produced, are extremely

sensitive to subtle variations in the driving laser's waveform, and it is not yet clear whether the observed link between high-order harmonic emission and single-band currents applies in materials other than silica. It also remains to be seen whether the laser-pulse parameters affect the mechanism of current production — although there is evidence suggesting that other mechanisms dominate when longer-wavelength lasers are used⁶.

Realization of light-wave-driven devices such as attosecond transistors, which could both switch and drive currents at multi-petahertz frequencies, will require a better understanding of the mechanisms that cause damage and heat accumulation in materials exposed to strong light fields, and of atomic-scale electron motion in solids — in particular how the electronic band structure of a material is modified in strong light fields. Nevertheless, the first hurdles have been cleared: Garg and colleagues' measurements show not only that multi-PHz currents can be reproducibly and controllably generated, but also that they can be measured in real time using attosecond optical techniques. ■

Michael Chini is in the Department of Physics, University of Central Florida, Orlando, Florida 32816-2385, USA.
e-mail: michael.chini@ucf.edu

1. Garg, M. *et al.* *Nature* **538**, 359–362 (2016).
2. Schiffrin, A. *et al.* *Nature* **493**, 70–74 (2013).
3. Paasch-Colberg, T. *et al.* *Nature Photon.* **8**, 214–218 (2014).
4. Ghimire, S. *et al.* *Nature Phys.* **7**, 138–141 (2011).
5. Luu, T. T. *et al.* *Nature* **521**, 498–502 (2015).
6. Vampa, G. *et al.* *Nature* **522**, 462–464 (2015).

CELL BIOLOGY

The organelle replication connection

Live-cell imaging reveals that a functional interaction occurs between two different organelles: contact between the endoplasmic reticulum and mitochondria is needed for mitochondrial DNA replication and division.

ELENA ZIVIANI & LUCA SCORRANO

The most fundamental difference between prokaryotic and eukaryotic cells is the presence of membrane-bounded organelles in eukaryotic cells. Organelles, such as mitochondria, chloroplasts and the endoplasmic reticulum, allow eukaryotic cells to form microenvironments in which biological processes can be spatially and temporally regulated¹. The nuclear genome encodes most organellar proteins, although certain organelles, such as mitochondria

and chloroplasts, contain some of their own genetic information. Coordination between the organellar genome and the nuclear genome is therefore required to ensure correct DNA content, DNA replication and protein translation. Writing in *Science*, Lewis *et al.*² investigate whether mitochondrial DNA is replicated at random or at specific locations within the cell, using a live-cell microscope-imaging approach to monitor mitochondrial DNA replication in human cells.

Organelles are enclosed by a lipid bilayer that forms their external boundary. The

bilayer is impermeable to most molecules — a prerequisite for the creation of functionally specialized spaces. The fundamental question of how organelles communicate with their external surroundings is still under investigation. The lipid bilayer of each organelle contains transport proteins that can allow the import and export of specific proteins and metabolites. However, cellular communication does not consist solely of soluble signals — cells can also sense their microenvironments through physical and mechanical cues³. This type of sensing might also apply to intracellular organelles, in which case the shape of juxtaposed organellar compartments could potentially affect organelle communication and fundamental biological processes within the cell.

Lewis and colleagues now provide insight into the relationship between organellar structure, the process of mitochondrial DNA synthesis and the transmission of the replicated mitochondrial DNA to daughter mitochondria. Sites where mitochondria are associated with the endoplasmic reticulum (ER) have previously been identified⁴ in yeast and mammalian cells as locations associated with mitochondrial division. Lewis *et al.* investigated mitochondrial DNA replication in living human cells using fluorescence microscopy techniques that enabled them to monitor the location of organelles and key intracellular components, including the mitochondrial DNA polymerase protein and a mitochondrial division enzyme. They found a link between the location of replicating nucleoids (the discrete units of mitochondrial DNA within the mitochondria) and the sites of contact between mitochondria and the tubular ER (Fig. 1a).

The authors reasoned that for nucleoids to be distributed equally into daughter mitochondria, nucleoid replication would have to occur at or close to the site of mitochondrial division. They reported that replication occurred close to the point of contact between the mitochondrion and the ER — an association that was originally described in yeast⁵.

Lewis and colleagues' study provides a leap forward in our understanding by also showing that ER structure impinges on the regulation of mitochondrial DNA homeostasis. The authors manipulated the levels of certain ER proteins to shift the ER structure from a tubular form to a sheet-like form, and they observed by microscopy that the number of nucleoids undergoing DNA replication was reduced, although there were no changes in the total mitochondrial DNA content (Fig. 1b). In our opinion, this work indicates for the first time that the shape of one organelle has a role in determining a key function of a different juxtaposed organelle. Yet, from this study it remains unclear whether the observed effects are directly mediated by a protein complex that

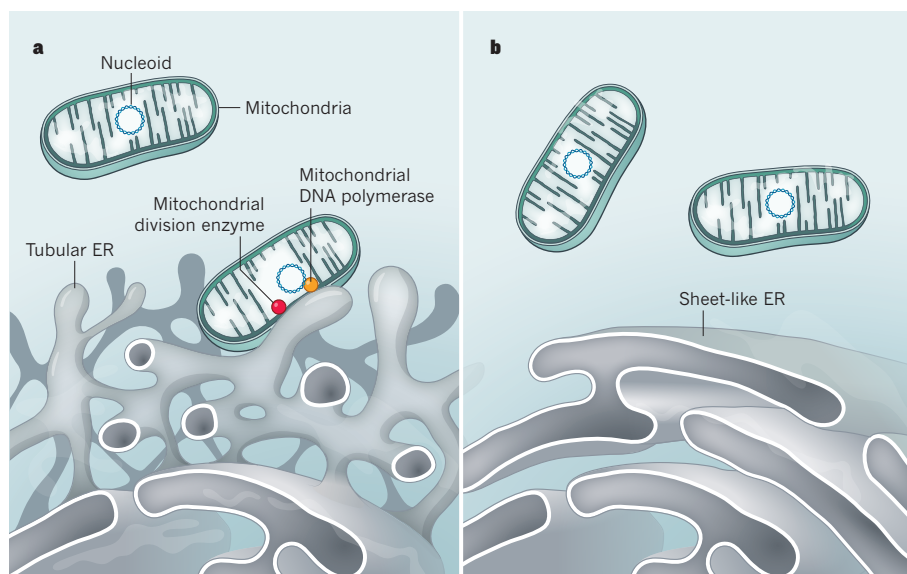


Figure 1 | The cellular location of mitochondrial DNA replication. Using microscopy in living human cells, Lewis *et al.*² monitored replication of the DNA structures called nucleoids in mitochondria. **a**, The authors observed that sites of DNA replication, identified from co-localization of nucleoids with either mitochondrial DNA polymerase protein or a mitochondrial division enzyme, were located in regions of the mitochondrion in close association with endoplasmic reticulum (ER) tubule structures. **b**, When the authors manipulated the ER so that it was mainly in a sheet-like form rather than in a tubular form, they observed a decrease in nucleoid replication.

connects the ER to the replicating mitochondrial DNA, or whether the effects are mediated indirectly by some unidentified messengers.

The notion that organelle shape can influence function is widespread in biology and is generally appreciated for mitochondria¹. However, there has been no previous hint that mitochondrial DNA maintenance and transmission, those key processes for mitochondrial function and cell survival, could be influenced by physical inputs at the interaction interface between mitochondria and the ER. As well as having relevance for mitochondrial diseases, perhaps more importantly, this work raises questions about the role of physical interactions in cross-talk between organelles.

Several questions remain open. For example, it is unclear how shape controls function, and how physical forces might be translated into biological responses. One possibility is that structural changes in the ER might promote remodelling of the cell's protein-filament network, called the cytoskeleton, which might in turn result in the recruitment of specific cellular components such as lipids or proteins to form specialized microdomains on the organelle's surface. Another possibility is that, depending on the sheet-like or tubular structure of the ER, the external forces and physical constraints generated might be sensed locally at the mitochondrion's outer surface to promote activation of a genetic program that affects mitochondrial DNA replication. Analyses of genes that are differentially expressed in association with changes in ER shape might pave

the way for studies to determine the mechanisms underlying the relationship between the ER contact and mitochondrial DNA replication.

It is unclear whether tethering molecules might be involved in this process and, if so, which molecules are responsible. Tethers between the ER and mitochondria exist in yeast⁶ and mammalian⁷ cells. Proving that mitochondrial DNA replication occurs at sites of organelle tethering would require genetic experiments to delete these tethering structures.

Lewis *et al.* have extended our understanding of the role of organelle cross-talk to include the control of mitochondrial DNA replication by the ER. Their work opens new avenues of research to explore how one organelle can have a profound influence on a neighbouring one. ■

Elena Ziviani is at the IRCCS, Ospedale San Camillo, 30126 Venice, Italy, and

Luca Scorrano is at the Venetian Institute of Molecular Medicine, 35129 Padua, Italy. They are also in the Department of Biology, University of Padua.

e-mails: elena.ziviani@unipd.it;
luca.scorrano@unipd.it

1. Pernas, L. & Scorrano, L. *Annu. Rev. Physiol.* **78**, 505–531 (2015).
2. Lewis, S. C., Uchiyama, L. F. & Nunnari, J. *Science* **353**, aaf5549 (2016).
3. Jaalouk, D. E. & Lammerding, J. *Nature Rev. Mol. Cell Biol.* **10**, 63–73 (2009).
4. Friedman, J. R. *et al. Science* **334**, 358–362 (2011).
5. Murley, A. *et al. eLife* **2**, e00422 (2013).
6. Kornmann, B. *et al. Science* **325**, 477–481 (2009).
7. de Brito, O. M. & Scorrano, L. *Nature* **456**, 605–610 (2008).

Accurate *de novo* design of hyperstable constrained peptides

Gaurav Bhardwaj^{1,2*}, Vikram Khipple Mulligan^{1,2*}, Christopher D. Bahl^{1,2*}, Jason M. Gilmore^{1,2}, Peta J. Harvey³, Olivier Cheneval³, Garry W. Buchko⁴, Surya V. S. R. K. Pulavarti⁵, Quentin Kaas³, Alexander Eletsy⁵, Po-Ssu Huang^{1,2}, William A. Johnsen⁶, Per Jr Greisen^{1,2,7}, Gabriel J. Rocklin^{1,2}, Yifan Song^{1,2,8}, Thomas W. Linsky^{1,2}, Andrew Watkins⁹, Stephen A. Rettie², Xianzhong Xu⁵, Lauren P. Carter², Richard Bonneau^{10,11}, James M. Olson⁶, Evangelos Coutsi¹², Colin E. Correnti⁶, Thomas Szyperski⁵, David J. Craik³ & David Baker^{1,2,13}

Naturally occurring, pharmacologically active peptides constrained with covalent crosslinks generally have shapes that have evolved to fit precisely into binding pockets on their targets. Such peptides can have excellent pharmaceutical properties, combining the stability and tissue penetration of small-molecule drugs with the specificity of much larger protein therapeutics. The ability to design constrained peptides with precisely specified tertiary structures would enable the design of shape-complementary inhibitors of arbitrary targets. Here we describe the development of computational methods for accurate *de novo* design of conformationally restricted peptides, and the use of these methods to design 18–47 residue, disulfide-crosslinked peptides, a subset of which are heterochiral and/or N–C backbone-cyclized. Both genetically encodable and non-canonical peptides are exceptionally stable to thermal and chemical denaturation, and 12 experimentally determined X-ray and NMR structures are nearly identical to the computational design models. The computational design methods and stable scaffolds presented here provide the basis for development of a new generation of peptide-based drugs.

The vast majority of drugs currently approved for use in humans are either proteins or small molecules. Lying between the two in size, and integrating the advantages of both^{1,2}, constrained peptides are an under-explored frontier for drug discovery. Naturally occurring constrained peptides, such as conotoxins, chlorotoxin, knottins and cyclotides, play critical roles in signalling, virulence and immunity, and are among the most potent pharmacologically active compounds known³. These peptides are constrained by disulfide bonds or backbone cyclization to favour binding-competent conformations that precisely complement their targets. Inspired by the potency of these compounds, there have been considerable efforts to generate new bioactive molecules by re-engineering existing constrained peptides using loop grafting, sequence randomization and selection⁴. Although powerful, these approaches are hindered by the limited variety of naturally occurring constrained peptide structures and the inability to achieve global shape complementarity with targets. There is need for a method of creating constrained peptides with new structures and functions that provides precise control over the size and shape of the designed molecules. A method with sufficient generality to incorporate non-canonical backbones and unnatural amino acids would enable access to broad regions of peptide structure and function space not explored by evolution.

Although there have been recent advances in protein design methodology^{5–9}, the computational design of covalently constrained peptides with new structures and non-canonical backbones presents new challenges. First, both backbone generation and design validation by structure prediction require new backbone sampling methods that can handle cyclic and mixed-chirality backbones. Second, methods are

needed for incorporation of multiple covalent geometric constraints without introduction of conformational strain. Third, energy evaluations must correctly model amino acid chirality.

Here we describe the development of new computational methods that meet these challenges, opening this frontier to computational design. We demonstrate the power of the methods by designing a structurally diverse array of 18–47 residue peptides spanning two broad categories: (i) genetically encodable disulfide-rich peptides, and (ii) heterochiral peptides with non-canonical sequences. Genetic encodability has the advantage of being compatible with high-throughput selection methods, such as phage, ribosome and yeast display, while incorporation of non-canonical components allows access to new types of structures, and can confer enhanced pharmacokinetic properties. To explore the folds accessible to genetically encoded constrained peptides under 50 amino acids, we selected nine topologies: HH, HHH, EHE, EEH, HEEE, EHEE, EEHE, EEEH and EEEEE (Fig. 1; we define a ‘topology’ as the sequence of secondary structure elements in the folded peptide, where H denotes α -helix and E denotes β -strand). To explore the expanded design space accessible with inclusion of non-canonical amino acids and backbone cyclization, we sought to cover topologies containing two to three canonical secondary structure elements: HH, HHH, EEH, EHE, HEE and EE, along with H_LH_R , a cyclic topology with left- and right-handed helices.

All of the design calculations described in this Article were carried out with the Rosetta software suite¹⁰ and followed the same basic approach. Large numbers of peptide backbones were stochastically generated as described in the following sections, combinatorial sequence design

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. ³Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. ⁴Seattle Structural Genomics Center for Infectious Diseases, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99352, USA. ⁵Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260, USA. ⁶Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ⁷Global Research, Novo Nordisk A/S, DK-2760 Måløv, Denmark. ⁸Cyrus Biotechnology, Seattle, Washington 98109, USA. ⁹Department of Chemistry, New York University, New York, New York 10003, USA. ¹⁰Department of Biology, New York University, New York, New York 10003, USA. ¹¹Center for Computational Biology, Simons Foundation, New York, New York 10010, USA. ¹²Applied Mathematics and Statistics and Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, USA. ¹³Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

*These authors contributed equally to this work.

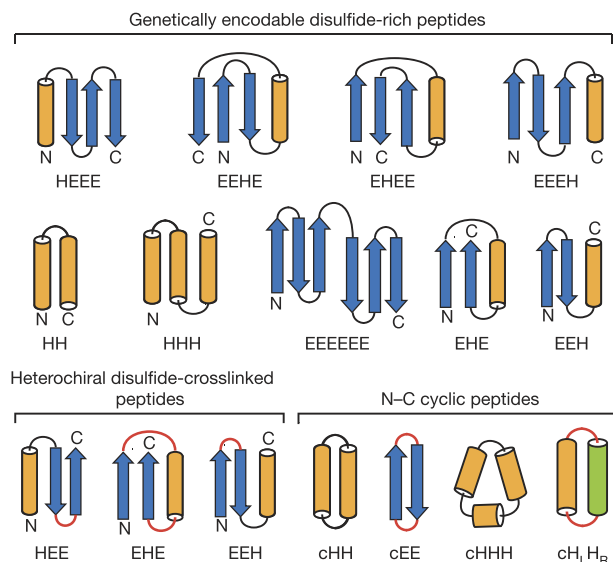


Figure 1 | Designed peptide topologies. The designed secondary structure architectures for each of the three classes of constrained peptides (genetically encodable disulfide-rich, heterochiral disulfide-crosslinked, and N-C cyclic) span most of the topologies that can be formed with four or fewer secondary structure elements. Arrows, β -strands; orange cylinders, right-handed α -helices; green cylinder, left-handed α -helix; red, loop segments containing D-amino acid residues.

calculations were carried out to identify sequences (including disulfide crosslinks) stabilizing each backbone conformation, and the designed sequence–structure pairs were assessed by determining the energy gap between the designed structure and alternative structures found in large-scale structure prediction calculations for the designed sequence. A subset of the designs in deep energy minima were then produced in the laboratory, and their stabilities and structures were determined experimentally.

Genetically encodable disulfide-constrained peptides

To design disulfide-stabilized genetically encodable peptides, we created a ‘blueprint’ specifying the lengths of each secondary structure element and connecting loop for each topology. Ensembles of backbone conformations were generated for each blueprint by Monte Carlo-based assembly of short protein fragments⁹, or, in the case of HH and HHH topologies, by varying the parameters in backbone generating equations¹¹. The backbones were scanned for sites capable of hosting disulfide bonds with near-ideal geometry, and one to three disulfide bonds were incorporated. Low-energy amino acid sequences were designed for each disulfide-crosslinked backbone using iterative rounds of Monte Carlo-based combinatorial sequence optimization while allowing the backbone and disulfide linkages to relax in the Rosetta all-atom force field (see Methods). Except for the EHEE topology, we performed no manual amino acid sequence optimization. Rosetta *ab initio* structure prediction calculations were carried out for each designed sequence, and synthetic genes were obtained for a diverse set of 130 designs for which the target structure was in a deep global free-energy minimum (Fig. 2a, b).

Disulfide bonds in peptides are unlikely to form in the reducing environment of the cytoplasm, so designs were secreted from *Escherichia coli* or cultured mammalian cells¹² (see Methods). Twenty-nine designs exhibited a redox-sensitive gel-shift, redox-sensitive high-performance liquid chromatography (HPLC) migration, and/or a circular dichroism (CD) spectrum consistent with the designed topology (see Supplementary Document 3). All 29 contain at least one non-alanine hydrophobic residue on each secondary structure element contributing van der Waals interactions in the core, which are probably important for proper peptide folding. We chose one representative design from each topology for further biochemical characterization.

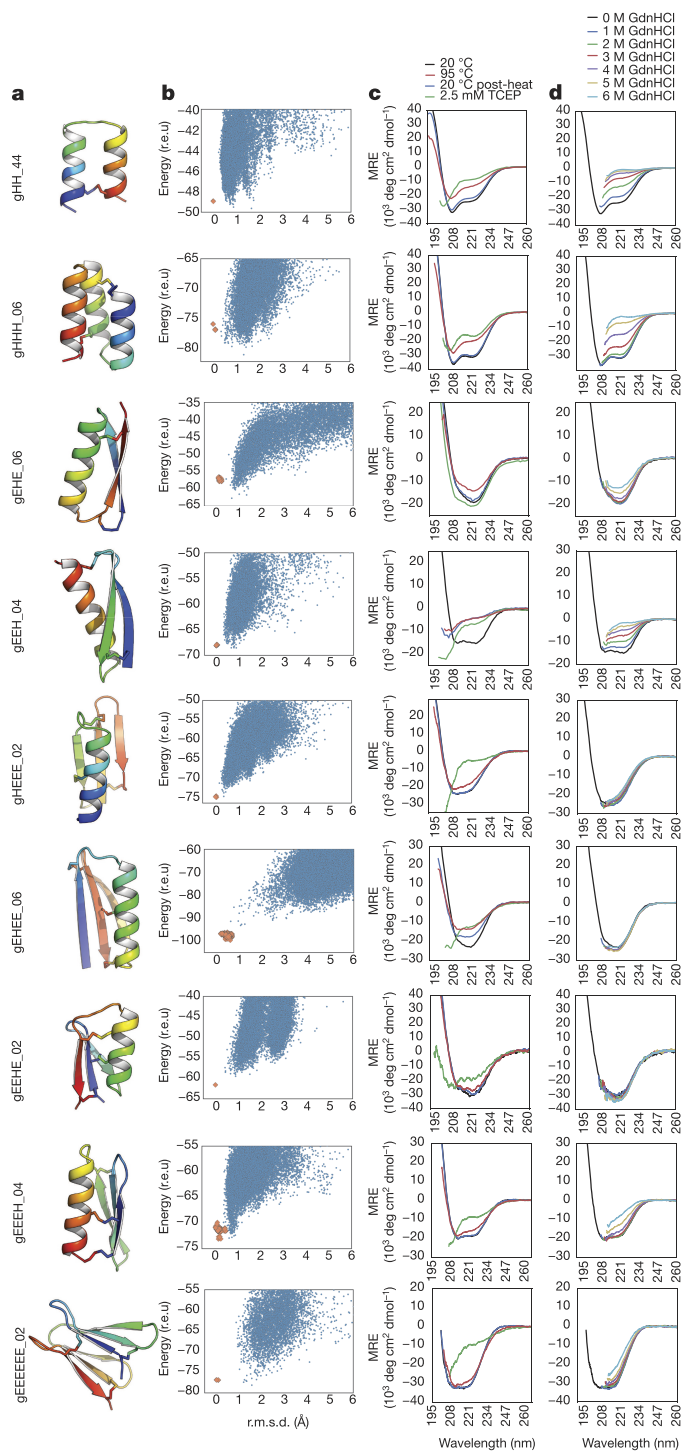


Figure 2 | Computational design and biophysical characterization of genetically encodable disulfide-rich peptides. Genetically encodable peptides are given the prefix ‘g’ and a number to differentiate designs that share a common topology (peptide name at far left).

a, Cartoon renderings of each design shown with rainbow colouring from the N terminus (blue) to the C terminus (red); disulfide bonds are shown as sticks. **b**, The energy landscape of each designed sequence was assessed by Rosetta structure prediction calculations starting from an extended chain (blue dots) or from the design model (orange dots); lower energy structures were sometimes sampled in the former because disulfide constraints were only present in the latter (r.e.u., Rosetta energy units; r.m.s.d., root mean square deviation from the designed topology). **c**, CD spectra at 20 °C (black lines), after heating to 95 °C (red lines), and upon cooling back to 20 °C (blue lines). Spectra collected with 2.5 mM TCEP are shown in green (MRE, mean residue ellipticity). **d**, CD spectra as a function of GdnHCl concentration (see key).

Since eight of the nine topologies contained four or more cysteine residues, we used multiple-stage mass spectrometry to investigate the disulfide connectivity. In all cases the data were consistent with the designed connectivity (see Supplementary Document 4).

The stability of the designs to thermal and chemical denaturation was assessed by CD spectroscopy. Samples were heated to 95 °C (Fig. 2c), or incubated with increasing concentrations of guanidinium hydrochloride (GdnHCl) (Fig. 2d). The contribution of disulfide bonds to protein folding was assessed by incubating samples with a ~100-fold molar excess of the reductant tris(2-carboxyethyl)phosphine (TCEP). Designs gHEEE_02, gEEEH_04 and gEEEEEE_02 are resistant to both thermal and chemical denaturation, while design gHH_44 is resistant to thermal denaturation. gHEEE_02 contains three disulfide bonds, with each secondary structure element participating in at least one disulfide bond, and no two secondary structure elements sharing more than one disulfide bond. gEEEH_04 has two of three disulfide bonds linking the N-terminal β -strand to the C-terminal α -helix. gEEEEEE_02 consists of two antiparallel β -sheets packing against one another in a sandwich-like arrangement, with each β -sheet stabilized by a disulfide bond linking one terminus to its adjacent β -strand. gHH_44 consists of two α -helices with a single disulfide bond connecting the termini.

We crystallized design gEHEE_06 and determined the structure to a resolution of 2.09 Å (Fig. 3, Supplementary Table 2-2). The crystals had three-fold non-crystallographic symmetry, and each protomer aligns to the design model with a mean all-atom root mean square deviation (r.m.s.d.) of 1.12 Å. All three of the designed disulfide bonds were well-defined by electron density (Extended Data Fig. 1), and rotamers of core residues exhibited excellent agreement with the design model. The protein was thermostable and completely resistant to chemical denaturation (Fig. 2c, d). While gEHEE_06 shares the short-chain scorpion toxin topology, the length of secondary structure elements and loops, and the position of the disulfide bonds, are entirely divergent from known natural peptides.

As crystallization efforts for other designs were unsuccessful (with phase-separation rather than protein precipitation observed), we expressed isotope-labelled peptides in *E. coli*, and determined structures by nuclear magnetic resonance (NMR) spectroscopy^{13,14} (see Methods). Upfield chemical shifts of the cysteine β -carbons¹⁵ (deposited in the Biological Magnetic Resonance Data Bank) confirmed the formation of the designed disulfide bonds. Design gEEHE_02, with one disulfide bond connecting the termini within the β -sheet and two between the α -helix and β -sheet, aligns to the NMR ensemble with a mean all-atom r.m.s.d. of 1.44 Å. This design was impervious to both thermal and chemical denaturation (monitored by CD spectroscopy), and remained partially folded in the presence of TCEP. The final three designs are each composed of three secondary structure elements, with termini located at opposite ends of the molecule and two disulfide bonds connecting each terminus to the middle structural element or adjacent loop. gEEH_04 was less resistant than the others to thermal denaturation, but its NMR structure is nearly identical to the design model (mean all-atom r.m.s.d., 1.29 Å). gEHE_06, which contains a solvent-exposed two-strand parallel β -sheet (rare in natural protein structures¹⁶), aligns to the NMR ensemble with an all-atom mean r.m.s.d. of 1.95 Å; it was thermally and chemically stable based on CD measurements, and remained folded in the presence of TCEP. gHHH_06 partially unfolds upon heating to 95 °C but returns to the folded state upon cooling; the design model aligns to the NMR ensemble with a mean all-atom r.m.s.d. of 1.74 Å. Taken together, the X-ray crystallographic and NMR structures demonstrate that our computational approach enables accurate design of protein main-chain conformation, disulfide bonds and core residue rotamers.

Synthetic heterochiral disulfide-constrained peptides

We next sought to design shorter disulfide-constrained peptides incorporating both L- and D-amino acids. We generalized the Rosetta energy

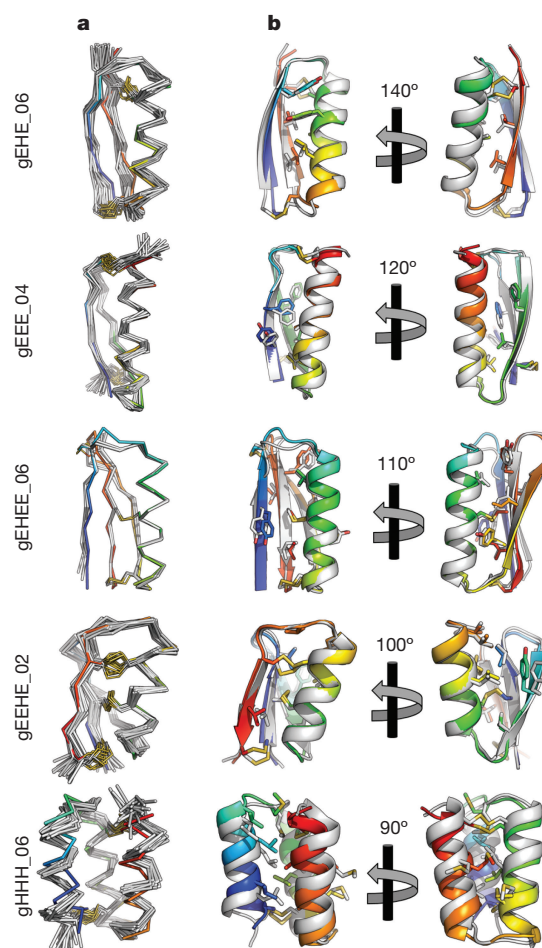


Figure 3 | X-ray crystal structures and NMR solution structures of designed peptides are very close to design models. Structures for gEHEE_06, gEEH_04, gEEHE_02 and gHHH_06 were determined by NMR spectroscopy, and the structure of gEHEE_06 was determined by X-ray crystallography. **a**, C_{α} traces of NMR ensembles, or superimposed members of the asymmetric unit, (grey), are aligned against the design model (rainbow). Disulfide bonds are shown with sidechain atoms rendered as sticks with sulfur atoms coloured yellow. **b**, Cartoon representation of the lowest energy conformer of each NMR ensemble or crystallographic asymmetric unit (grey) is shown aligned to the design model (rainbow). Two views of each structure are shown, rotated about the vertical axis by the indicated amount. Sidechain atoms of hydrophobic core residues are rendered as sticks.

function to support D-amino acids by inverting the torsional potentials used for the equivalent L-amino acids (see Methods and Supplementary Information), and sequence design algorithms were extended to enable mixed-chirality design. Since chemical synthesis is labour-intensive, we prioritized the development of automated computational screening techniques, supplementing Rosetta *ab initio* screening with molecular dynamics (MD) evaluation.

Large numbers of disulfide-constrained backbones for topologies HEE, EHE and EEH were generated by fragment assembly as described above for genetically encodable peptides. Sequences were designed (favouring D-amino acids at positions with positive mainchain ϕ dihedral angle values), and the resultant low-energy designs were evaluated using MD and *ab initio* structure prediction (Extended Data Fig. 2). For each topology, we selected a single, low-energy design (Extended Data Fig. 3) which underwent only small (<1.0 Å r.m.s.d.) fluctuations in the MD simulations (Extended Data Fig. 4) and had a large energy gap in the structure prediction calculations. Selected peptides were chemically synthesized, and structurally characterized by NMR. In all three cases, the NMR spectra had well-dispersed, sharp

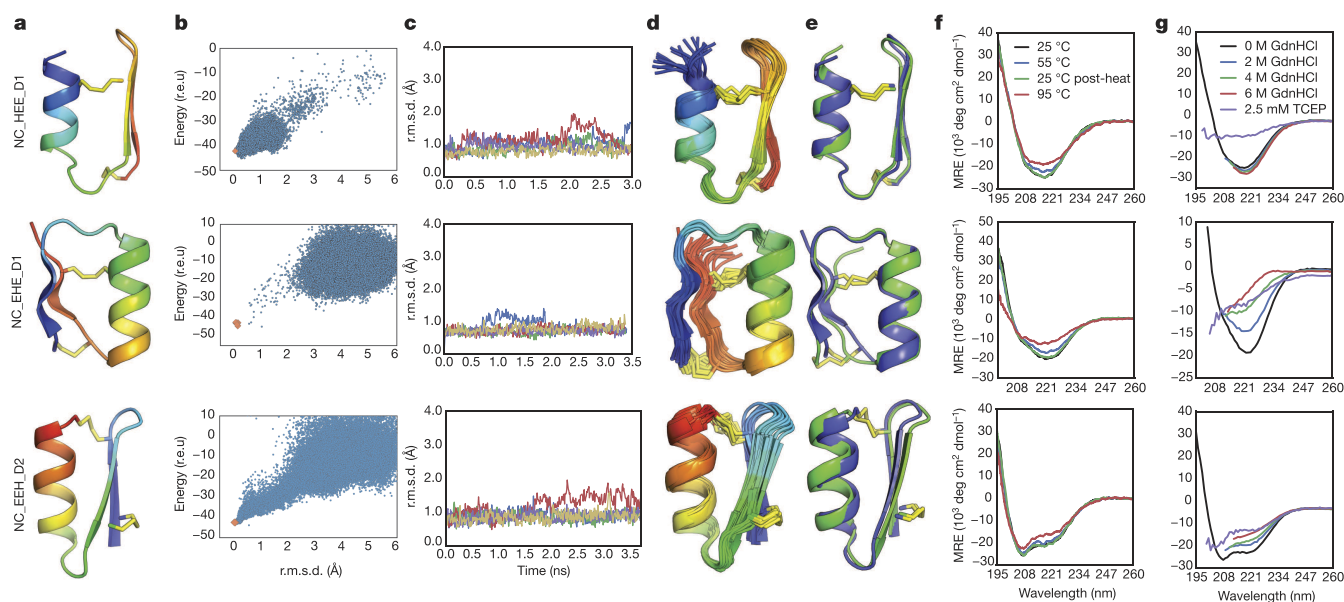


Figure 4 | Design and characterization of heterochiral disulfide-constrained peptides. The prefix ‘NC’ denotes non-canonical sequence or backbone architecture, and a numerical suffix differentiates designs sharing a common topology. **a**, Cartoon representations of design models with the N terminus in blue and C terminus in red. **b**, Folding energy landscapes from Rosetta *ab initio* structure prediction calculations. Blue dots indicate lowest-energy structures identified in independent Monte Carlo trajectories. Orange dots are from trajectories starting with the design model. (r.e.u., Rosetta energy units; r.m.s.d., root mean square deviation from the designed topology). **c**, Five representative trajectories from a total of 50 independent MD simulations starting from the design

peaks and secondary α proton ($^1\text{H}_\alpha$) chemical shifts consistent with the secondary structure of the design model (Supplementary Fig. 2–5).

High-resolution NMR solution structures were determined for each of the designs (Supplementary Table 2–3). NC_HEE_D1 is a 27-residue peptide with a D-proline, L-proline turn at the β – β junction; in this case, Rosetta re-identified a motif known previously to stabilize type II’ turns^{17,18}. The NMR structure closely matches the design model: the r.m.s.d. over all mainchain α carbon atoms (C_α r.m.s.d.) is 0.99 Å between the designed structure and the lowest-energy NMR model (Fig. 4, top row). NC_EHE_D1 is a 26-residue peptide crosslinked using two disulfide bonds with a D-arginine residue in the β – α loop and a D-asparagine residue as the C-terminal capping residue for the α -helix. The design model has a 1.9 Å C_α r.m.s.d. to the lowest-energy NMR ensemble member, and a 0.68 Å C_α r.m.s.d. to the closest member of the ensemble (Fig. 4, middle row); the last two residues at the C-terminal vary considerably in the ensemble). NMR characterization of the NC_EEH_D1 design showed an unwound C-terminal α -helix adopting an extended conformation, differing from the design model (Extended Data Fig. 5). We hypothesized that substantial strain was introduced by the angle between the helix and the preceding strand, and by the disulfide bonds at both ends of the helix. A second design for the same topology, NC_EEH_D2, has a type I’ turn at the β – β connection and a different disulfide pattern. The NMR ensemble for NC_EEH_D2 is very close to the design model (0.86 Å C_α r.m.s.d. to the lowest-energy NMR model; Fig. 4, bottom row).

We explored the stability of the designed peptides using CD spectroscopy to monitor thermal and chemical denaturation. All three peptides are very thermostable; there is no loss in secondary structure for NC_HEE_D1 and NC_EEH_D2 at 95 °C, and only a small decrease for NC_EHE_D1 (Fig. 4f). Remarkably, NC_HEE_D1 does not denature in 6 M GdnHCl (Fig. 4g, top row). Treatment with TCEP causes unfolding of all three designs, highlighting the importance of disulfide bonds.

All of the designs described in this Article were created *de novo* without sequence information from natural proteins. Searches for

model with different initial velocities. **d**, NMR-determined structure ensembles. Cartoon representations coloured and oriented as in **a**, **e**. Superposition of the designed structure (blue) with the lowest-energy NMR structure (green). **f**, CD wavelength spectra between 195 nm and 260 nm recorded at 25 °C (black), 55 °C (blue), 95 °C (red) and after cooling back to 25 °C (green). **g**, CD spectra recorded at 0 M (black), 2 M (blue), 4 M (green), or 6 M GdnHCl (red), or with 2.5 mM TCEP/0 M GdnHCl (purple). Data are truncated in the far-ultraviolet region for spectra acquired in the presence of high GdnHCl concentrations (due to GdnHCl absorbance).

similar sequences in the PDB and NCBI NR database using PSI-BLAST found significant alignments (*e*-value < 0.01) only for NC_EHE_D1 and gHH_44 (Supplementary Table S1–2 and S1–3). The NC_EHE_D1 sequence has weak similarity (*e*-value of 2×10^{-4}) to the zinc-finger domain of lysine-specific demethylase (PDB ID: 2MA5), but the aligned regions adopt different structures (Extended Data Fig. 6). The gHH_44 sequence has weak similarity (*e*-value of 0.001) to a single long helix in a leucine zipper (PDB ID: 4R4L), very different from the helical hairpin topology of the design.

Synthetic backbone-cyclized peptides

Next, we explored the design of peptides with cyclized backbones, which can increase stability and protect against exopeptidases¹⁹. To generate such backbones without dependence on fragments of known structures, we implemented a generalized kinematic loop closure^{20,21} method (named ‘GenKIC’) to sample arbitrary covalently linked atom chains capable of connecting the termini. Each GenKIC chain-closure attempt involves perturbing multiple chain degrees of freedom, then analytically solving kinematic equations to enforce loop closure with ideal peptide bond geometry in the case of N–C cyclic peptides (see Methods, Supplementary Information, and Extended Data Fig. 7). Sequence design, backbone relaxation, and *in silico* structure validation using MD simulation and Rosetta *ab initio* structure prediction were carried out with terminal bond geometry constraints (Extended Data Fig. 2).

We synthesized cyclic peptides for three topologies (cEE, cHH and cHHH) and determined their structures by NMR spectroscopy. The 18-residue NC_cEE_D1 design has the cyclic anti-parallel β -sheet fold of natural θ -defensins, but with one disulfide bond (rather than three), and different turn types containing heterochiral sequences²². The lowest-energy NMR model has a C_α r.m.s.d. of 1.26 Å to the designed structure. The variability in the curvature of the sheets across the NMR ensemble is similar to the variability observed in the structure prediction calculations (Fig. 5, top row). The 26-residue NC_cHH_D1

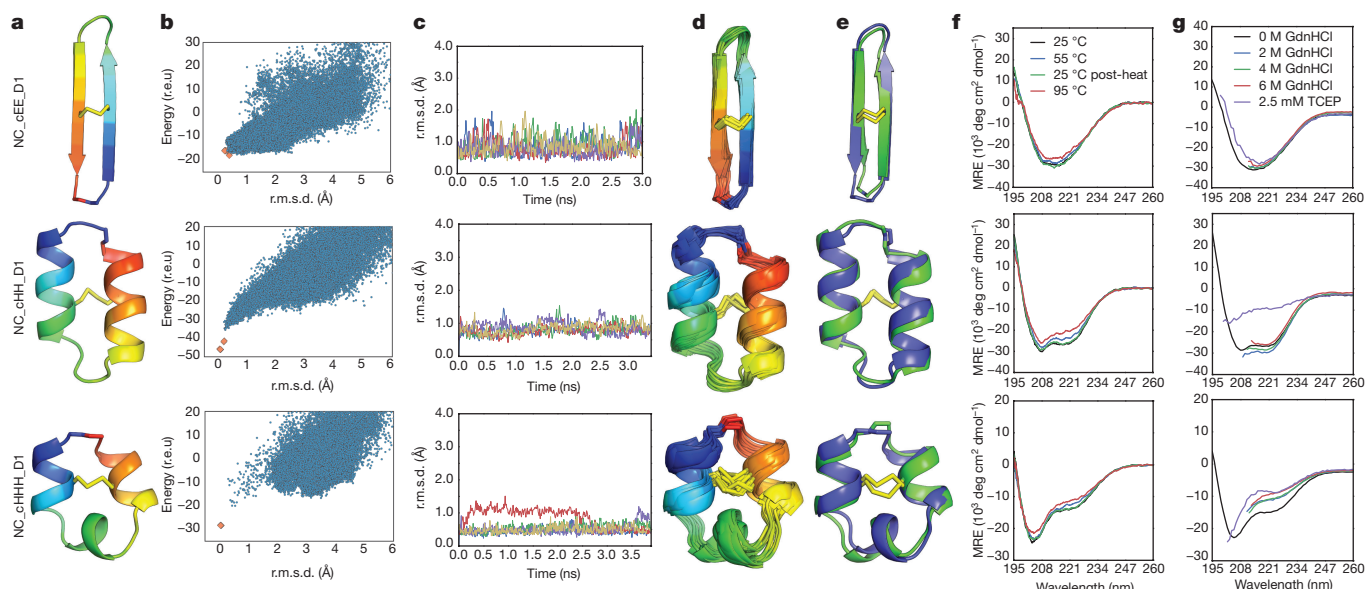


Figure 5 | Design and characterization of N-C backbone cyclic peptides. Peptide names at far left; columns a–g as in Fig. 4. A lower-case ‘c’ in the peptide name indicates N-C cyclic backbone.

design, which has one disulfide bond linking the two α -helices, has a 1.03 Å C_α r.m.s.d. from the lowest-energy NMR structure (Fig. 5, second row). The 22-residue NC_cHHH_D1 design has three short regions of α -helical structure and a single disulfide bond. The NMR structure of the design was again very close to the design model (Fig. 5, third row), with a C_α r.m.s.d. of 1.06 Å to the lowest-energy NMR structure.

All three cyclic topologies were found to be extremely stable in thermal denaturation experiments, retaining CD signal when heated to 95 °C (Fig. 5f). The CD spectra of NC_cHH_D1 and NC_cEE_D1 were nearly identical in 0 and 6 M GdnHCl, indicating that these peptides do not chemically denature (Fig. 5g; NC_cHHH_D1 showed some loss of secondary structure in 6 M GdnHCl). After treatment with TCEP, both NC_cHH_D1 and NC_cHHH_D1 lost secondary structure, but the CD spectrum of NC_cEE_D1 was not changed by reduction of the central disulfide bond (Fig. 5g, top row). Overall, the cyclic designs are exceptionally stable given their very small sizes.

Beyond natural secondary and tertiary structure

As a final test of the generality of the new design methodology, we designed a heterochiral, backbone-cyclized, two-helix topology with one non-canonical left-handed α -helix and one canonical right-handed α -helix (H_LH_R) assembling into a tertiary structure not observed in natural proteins. As before, we validated designs by MD; however, for validation by *ab initio* structure prediction it was necessary to develop a new, GenKIC-based structure prediction protocol (see Extended Data Fig. 8, Methods, and Supplementary Information) since the standard Rosetta *ab initio* structure prediction method utilizes fragments of native proteins, which typically do not contain left-handed helices. Our selected design for this topology, NC_ H_LH_R _D1, is a 26-residue peptide with one D-cysteine, L-cysteine disulfide bond connecting the right-handed and left-handed α -helices. There is an excellent match between the NMR structure ensemble and design model (C_α r.m.s.d., 0.79 Å) (Fig. 6). As expected for the nearly achiral topology, the CD signal is very small (as observed for a previously studied two-chain, four-helix mixed D/L system²³), and no change was observable on heating to 95 °C. The secondary $^1H_\alpha$ chemical shifts also show nearly no change on heating to 75 °C (Fig. 6g, Supplementary Fig. 2–6), indicating that the peptide is thermostable. Successful design of this topology demonstrates that our computational methods are sufficiently versatile and robust to design in a conformational space not explored by nature.

Conclusions

The key advances in computational design presented here—notably the methods for designing constrained peptide backbones spanning a broad range of topologies and incorporating natural and non-natural building-blocks—enable high-accuracy design of new peptides with exceptional thermostability and resistance to chemical denaturation. All 12 experimentally determined structures are in close agreement with the design models, including one with helices of different chirality. Unlike the natural constrained peptide families, designed peptides are not limited to particular shapes, sizes, nucleating motifs, or disulfide connectivities; indeed, the sequences of these *de novo* peptides are quite different from those of any known peptides. Here we have focused on extending sampling and scoring methods to permit design with D-amino acids and cyclic backbones, but the new tools are fully generalizable to peptides containing more exotic building-blocks, such as amino acids with non-canonical sidechains²⁴ or non-canonical backbones²⁵.

The hyperstable molecules presented in this study provide robust starting scaffolds for generating peptides that bind targets of interest using computational interface design²⁶ or experimental selection methods. Solvent-exposed hydrophobic residues can be introduced without impairing folding or solubility (Extended Data Figs 9 and 10, Supplementary Fig. 2–6), suggesting high mutational tolerance. Hence it should be possible to re-engineer the peptide surfaces, incorporating target-binding residues to construct binders, agonists or inhibitors. There has been considerable effort in both academia and industry to use small, naturally occurring proteins as alternatives to antibody scaffolds for library selection-based affinity reagent generation. Our genetically encoded designs offer considerable advantages as starting points for such approaches because of their high stability, small size and diverse shapes. Furthermore, having been designed exclusively to be robust and stable, they lack the often-destabilizing non-ideal structural features that arise in naturally occurring proteins from evolutionary selective pressure for a particular function. Similarly, the heterochiral designs described here provide starting points for split-pool and other selection strategies compatible with non-canonical amino acids.

Going beyond the re-engineering of our hyperstable designs to bind targets of interest, the methods developed in this Article can be used to design new backbones to fit specifically into target binding pockets. Such ‘on-demand’ target-specific scaffold generation is likely to yield scaffolds with considerably greater shape-complementarity than that of scaffolds generated without knowledge of the target. More generally,

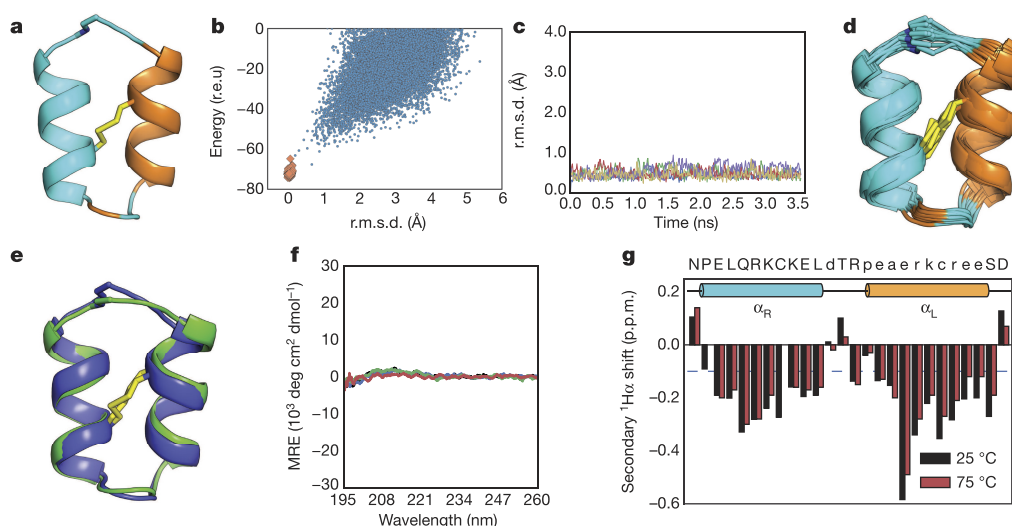


Figure 6 | Design and characterization of a peptide with non-canonical secondary and tertiary structure. **a**, NC_{H1HR}_D1 design (cyan, L-amino acids; orange, D-amino acids). **b**, Folding energy landscape generated using a new structure prediction algorithm compatible with non-canonical secondary structures (see Methods and Supplementary Information). **c**, Five representative MD trajectories (from a total of 50) starting from the design model with different initial velocities. **d**, NMR-determined structure ensembles, coloured and oriented as in **a**. **e**, Superposition of designed structure (blue) with lowest-energy NMR structure (green). **f**, CD spectra between 195 nm and 260 nm recorded

our computational methods open up previously inaccessible regions of shape space, and, in combination with computational interface design, should help unlock the pharmacological potential of peptide-based therapeutics.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 April; accepted 18 August 2016.

Published online 14 September 2016.

- Conibear, A. C. *et al.* Approaches to the stabilization of bioactive epitopes by grafting and peptide cyclization. *Biopolymers* **106**, 89–100 (2016).
- Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* **81**, 136–147 (2013).
- Góngora-Benítez, M., Tulla-Puche, J. & Albericio, F. Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem. Rev.* **114**, 901–926 (2014).
- Kimura, R. H., Levin, A. M., Cochran, F. V. & Cochran, J. R. Engineered cystine knot peptides that bind $\alpha_v\beta_3$, $\alpha_v\beta_5$, and $\alpha_5\beta_1$ integrins with low-nanomolar affinity. *Proteins* **77**, 359–369 (2009).
- Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
- Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
- Lin, Y.-R. *et al.* Control over overall shape and size in de novo designed proteins. *Proc. Natl Acad. Sci. USA* **112**, E5478–E5485 (2015).
- Doyle, L. *et al.* Rational design of α -helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
- Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
- Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
- Bandaranayake, A. D. *et al.* Daedalus: a robust, turnkey platform for rapid production of decigram quantities of active recombinant proteins in human cell lines using novel lentiviral vectors. *Nucleic Acids Res.* **39**, e143 (2011).
- Sagaram, U. S. *et al.* Structural and functional studies of a phosphatidic acid-binding antifungal plant defensin MtDef4: identification of an RGFRRL motif governing fungal cell entry. *PLoS One* **8**, e82485 (2013).

at 25 °C (black), 55 °C (blue), 95 °C (red) and after cooling back to 25 °C (green). The CD spectrum of NC_{H1HR}_D1 exhibits very weak signals because the L- and D- helical signals largely cancel. **g**, Secondary $^1\text{H}_\alpha$ chemical shifts (p.p.m.) are nearly identical from 25 °C (black) to 75 °C (red). NC_{H1HR}_D1 sequence displayed on top; orange cylinder, left-handed helix; cyan cylinder, right-handed helix; blue dashed line represents 0.1 p.p.m. of secondary $^1\text{H}_\alpha$ chemical shifts (groups of residues with secondary $^1\text{H}_\alpha$ shifts < -0.1 p.p.m. are typically indicative of helical regions).

- Liu, G. *et al.* NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc. Natl Acad. Sci. USA* **102**, 10487–10492 (2005).
- Sharma, D. & Rajarathnam, K. ^{13}C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* **18**, 165–171 (2000).
- Richardson, J. S. β -Sheet topology and the relatedness of proteins. *Nature* **268**, 495–500 (1977).
- Syud, F. A., Stanger, H. E. & Gellman, S. H. Interstrand side chain–side chain interactions in a designed β -hairpin: significance of both lateral and diagonal pairings. *J. Am. Chem. Soc.* **123**, 8667–8677 (2001).
- Lai, J. R., Huck, B. R., Weisblum, B. & Gellman, S. H. Design of non-cysteine-containing antimicrobial β -hairpins: structure-activity relationship studies with linear protegrin-1 analogues. *Biochemistry* **41**, 12835–12842 (2002).
- Wang, J., Yadav, V., Smart, A. L., Tajiri, S. & Basit, A. W. Toward oral delivery of biopharmaceuticals: an assessment of the gastrointestinal stability of 17 peptide drugs. *Mol. Pharm.* **12**, 966–973 (2015).
- Coutsias, E. A., Seok, C., Jacobson, M. P. & Dill, K. A. A kinematic view of loop closure. *J. Comput. Chem.* **25**, 510–528 (2004).
- Mandell, D. J., Coutsias, E. A. & Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6**, 551–552 (2009).
- Trabi, M., Schirra, H. J. & Craik, D. J. Three-dimensional structure of RTD-1, a cyclic antimicrobial defensin from Rhesus macaque leukocytes. *Biochemistry* **40**, 4211–4221 (2001).
- Sia, S. K. & Kim, P. S. A designed protein with packing between left-handed and right-handed helices. *Biochemistry* **40**, 8981–8989 (2001).
- Renfrew, P. D., Douglas Renfrew, P., Choi, E. J., Richard, B. & Brian, K. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS One* **7**, e32637 (2012).
- Drew, K. *et al.* Adding diverse noncanonical backbones to Rosetta: enabling peptidomimetic design. *PLoS One* **8**, e67051 (2013).
- Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements Computer time was awarded by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, a Department of Energy (DOE) Office of Science User Facility supported under contract DE-AC02-06CH11357. We thank the University of Washington Hyak supercomputing network for computing and data storage resources, and Rosetta@Home volunteer participants on BOINC for additional computing resources. We are grateful for facility access at the Queensland NMR Network. We thank D. Alonso, J. Bardwell, G. Bhabha, T.J. Brunette, D. Ekiert, A. Ford, N. Hasle, B. Keir, N. Koga, Y. Liu, D. Madden, B. Mao, D. May, V. Ovchinnikov,

S. Srivatsan, L. Stewart, R. van Deursen, and M. Williamson for help and advice, and R. Krishnamurthy, P. Hosseinzadeh, and A. Vorobieva for critical comments and manuscript suggestions. This work was supported by NIH grant P50 AG005136 supporting the Alzheimer's Disease Research Center, philanthropic gifts from the Three Dreamers and Washington Research Foundation, and funding from the Howard Hughes Medical Institute. The Australian Research Council funds D.J.C. as an Australian Laureate Fellow (FL150100146). C.D.B. was supported by NIH grant T32-H600035. T.S. acknowledges NIH support (GM094597), and S.V.S.R.K.P., A.E. and X.X. were supported with NESG funds. E.C. is funded by NIGMS GM090205. We thank P. Rupert and R.K. Strong at the Fred Hutchinson Cancer Research Center for aid in collecting and refining X-ray data for gEHEE_06. G.W.B. was funded by the National Institute of Allergy and Infectious Diseases, National Institute of Health, Department of Health and Human Services (Federal contract HHSN272201200025C). A portion of this research was performed using EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

Author Contributions C.D.B., G.B., V.K.M. and D.B. designed the study. V.K.M. developed algorithms with help from A.W., E.C., Y.S., G.B., R.B., C.D.B., G.J.R. and T.W.L. C.D.B. and J.M.G. designed canonical peptides with help from D.B., G.J.R. and T.W.L. G.B. designed heterochiral and backbone-cyclized

peptides with help from V.K.M., D.B., P.G. and P.S.H. C.D.B. expressed and characterized designed canonical peptides from *E. coli* with help from J.M.G. and S.A.R. J.M.G. performed MS analysis. W.A.G. and C.E.C. purified canonical peptides *via* Daedalus and determined X-ray crystal structures. G.W.B., S.V.S.R.K.P., A.E. and T.S. determined NMR solution structures of canonical peptides, purified with isotopic labelling by C.D.B. O.C. and G.B. synthesized, purified and characterized designed non-canonical peptides. P.J.H. and D.J.C. determined NMR solution structures of non-canonical peptides. P.J.H., Q.K. and D.J.C. analysed data from structure determination of non-canonical peptides. C.D.B., G.B., V.K.M. and D.B. wrote the manuscript with help from all authors.

Author Information Peptide structures have been deposited in the RCSB Protein Data Bank with accession codes 5JG9, 2ND2, 2ND3, 5JH1, 5J14, 5KVN, 5KWO, 5KWP, 5KWY, 5KX2, 5KWZ, 5KX1, 5KX0. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (dabaker@uw.edu).

Reviewer Information *Nature* thanks V. Nanda and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Computational design. *De novo* design of constrained peptides can be divided into two main steps: backbone assembly and sequence design. Practically, our peptide design pipeline has been optimized to permit these two steps to be performed in immediate succession with a single set of inputs, with no need for export or manual curation of the generated backbones before the sequence design. (A third and final validation step is typically performed separately.)

For backbone assembly, we used two different approaches: disulfide-constrained topologies were sampled using a fragment assembly method, whereas backbone-cyclized peptide topologies were sampled using a fragment-independent kinematic closure-driven approach. Example scripts and command lines for each step in the design workflow are available in Supplementary Information.

Backbone design using fragment assembly. In the case of disulfide-crosslinked designs, the topology was defined using a 'blueprint' that specifies secondary structure and torsion bins for each amino acid residue, the latter defined using the ABEGO alphabet system^{7,9}. The ABEGO nomenclature assigns a letter to each of five regions, or bins, in Ramachandran space. These bins correspond to the α -helical region (A), the β -sheet region (B), the region with positive mainchain ϕ dihedral angle values typically accessed by glycine (G), and the remainder of the Ramachandran space (E). (The fifth bin, O, represents residues with *cis*-peptide bonds, and was not used here.) The blueprint is the input for a Rosetta Monte Carlo-based fragment assembly protocol^{7,9,10,27} that generates backbone conformations that match the blueprint architecture. Briefly, the fragment assembly protocol uses the defined blueprint to pick backbone fragments from a database of non-redundant high-resolution crystal structures. The insertion of fragments serves as the moves in a Monte Carlo search of backbone conformation space. For searches of the NC_EEH topology, loop types were limited to ABEGO bins EA and GG for the $\beta\beta$ connection, and BAB and GBB for the $\alpha\beta$ connection. For sampling of the NC_EHE topology, $\beta\alpha$ connections were limited to GBB, BAB and AB, and $\alpha\beta$ connections were limited to GB, GBA and AGB. For sampling of the NC_HEE topology, $\alpha\beta$ connections were limited to BAAB, GB, GBA and AGB, and $\beta\beta$ connections were limited to EA and GG.

Backbone design using generalized kinematic closure. Although the fragment-based approaches described above are powerful, they are limited to conformations that are favoured by peptides composed primarily of L-amino acids. For N-C cyclic designs—NC_CHHH_D1, NC_CHH_D1, NC_CEE_D1 and NC_CH₁H_R_D1—we chose to focus on fragment-independent methods that are better suited for exploring conformations that are accessible to only mixed D/L peptides. We therefore turned to generalized kinematic closure (GenKIC).

GenKIC-based sampling works by treating a peptide as a loop, or a series of loops to be 'closed'. The torsion values of an initial, 'anchor' residue are randomly selected; this residue is then fixed, and the rest of the peptide is treated as a loop-closure problem. The particular covalent linkages serve as a set of geometric constraints for loop closure. The GenKIC algorithm performs a series of user-controlled perturbations to the torsion angles of the peptide chain, which inevitably disrupt the geometry of the closure points. GenKIC then mathematically solves for the value of six 'pivot' torsion angles that restore the geometry of the closure points and permit the loop to remain closed^{20,21,28}. Because the algorithm can return up to sixteen solutions per closure attempt, filters are applied to eliminate solutions with pivot amino acid residues in energetically unfavourable regions of Ramachandran space or with other geometric problems, such as clashes with other residues. The 'best' solution is then chosen on the basis of the Rosetta score function¹⁰.

During the sampling steps, regions in the design topology that were intended to form helices or sheets were initialized to ideal mainchain ϕ and ψ dihedral values, and were either kept fixed or perturbed by only small amounts ($<20^\circ$). In loop regions, the perturbation was carried out by drawing torsion values randomly, biased by the Ramachandran preferences of the amino acid residue. Glycine or D/L alanine were used for backbone sampling before design. The allowed range of the torsion value either covered the entire Ramachandran space or, in cases in which known loop ABEGO patterns could connect secondary structure elements, the mainchain torsion values were limited to those ABEGO bins. For example, during the design of the cEE topology, connection types were limited to the GG and EA torsion bins for the two-residue loops.

Disulfide positioning. To design disulfide bonds, we evaluated all of the residue pairs with C β atoms separated by ≤ 5 Å for geometry suitable to disulfide bond formation²⁷, selected backbones that could harbour disulfide bonds with near-ideal geometry, and incorporated one to three disulfide bonds. To select an ideal disulfide configuration from the set of all sterically possible combinations of disulfide bonds for a given backbone, we ranked disulfide configurations according to their effect on the configurational entropy of the unfolded state. The reduction in the entropy of the unfolded state due to a set of multiple crosslinks was computed according to a random flight model using equation (6) in ref. 29, with the volume

of tolerance, ΔV , equal to 29.65 \AA^3 and the link length, b , equal to 3.8 \AA . This method has been implemented in the Rosetta software suite as DisulfidizeMover and DisulfideEntropyFilter, both of which are accessible to the RosettaScripts scripting language.

Modifications to Rosetta to permit design of cyclic backbones and mixed D/L peptides. D-amino acid residues allow access to regions of conformational space that are normally accessed by only glycine. When placed correctly, they can provide greater rigidity than glycine, stabilizing glycine-dependent structural motifs and, thereby, the overall fold³⁰. Because the Rosetta software suite has primarily been used for designing proteins consisting of the 19 canonical L-amino acids and glycine, a number of modifications were necessary to permit robust design of peptides containing mixtures of D- and L-amino acids. First, Rosetta's default scoring function (*talaris2013* at the time of the work described here) was updated to permit D-amino acids to be scored with mirror symmetry relative to their L-counterparts. Terms in the score function that are based on mainchain or sidechain torsion values were modified to invert D-amino acid torsion values before applying the equivalent L-amino acid potentials. The score-function terms that are based on interatomic distances required minimal changes. To permit energy minimization, score-function derivatives were also modified to invert torsion derivative values for D-amino acids. Rosetta's rotameric search algorithm, the *packer*, was modified to use L-amino acid rotamers with sidechain χ torsion values inverted for D-amino acid rotamer packing, and to update H α and C β positions appropriately when inverting residue chirality. Finally, we added an option to symmetrize the energy tables for the mainchain torsion preferences of glycine, which are asymmetric by default because they are based on statistics taken from the Protein Data Bank (PDB). (Glycine, in the context of L-amino acids only, occurs disproportionately in the positive- ϕ region of Ramachandran space, but should have no asymmetric preferences in a mixed D/L context.) Details of these modifications are described in Supplementary Information.

Because Rosetta has traditionally been used to build linear polymers, a number of core Rosetta libraries had to be modified to permit N-C cyclic geometry to be sampled and scored properly. The assumption that residue i is connected to residues $i + 1$ and $i - 1$, which is invalid for cyclic peptides, has been removed and replaced with proper look-ups of connected residue indices. Cyclic geometry support was tested by confirming that the circular permutations of cyclic peptide models score identically.

As of 11 March 2016, the default Rosetta score function has been changed to *talaris2014*, which re-weights a number of score terms and introduces one new term³¹. The *talaris2014* score function has also been made fully compatible with D-amino acids and cyclic geometry. A newer, experimental score function, currently called *beta_nov15*, has also been made fully compatible with D-amino acids and cyclic geometry.

Sequence design and filtering. Backbone assembly using fragment assembly or GenKIC was followed by a sequence design step. Sequence design was performed using the FastDesign protocol (see Supplementary Information). This involves four rounds of alternating sidechain rotamer optimization (during which sidechain identities were permitted to change) and gradient-descent-based energy minimization. The best-scoring structure was taken from a minimum of three repeats of FastDesign (twelve rounds of rotamer optimization and minimization). Each amino acid position was sorted into a layer ('core', 'boundary' or 'surface') on the basis of burial, and the layer dictated the possible amino acid types allowed at that position; for example, hydrophobic amino acid residues were only permitted at core positions. To favour more proline residues during sequence design, the reference weight for proline in the Rosetta score function was reduced by 0.5 units. Backbones were allowed to move during the relaxation steps. For each topology, about 80,000 structures were generated, and filtered on the basis of the overall energy per residue, score terms related to backbone quality and score terms related to the disulfide geometry. In a few cases for non-canonical peptides, a conservative mutation was manually introduced into a surface-exposed repeat sequence (for example, an arginine to break a poly-lysine sequence) to facilitate unambiguous NMR assignment.

Rosetta-based computational validation. Typically, the number of designs that can be created *in silico* exceeds the number that can be produced and examined experimentally. We therefore used Rosetta to prune the list of designs, by one of two methods. For designs consisting of canonical amino acids, Rosetta's fragment-based *ab initio* algorithm³² was used to predict the structure of a design given its amino acid sequence, and to determine whether the target structure was a unique minimum in the conformational energy landscape. Disulfide bonds were not allowed to form during these simulations; the designed disulfide bonds are intended to stabilize an already unique global energy minimum, rather than to create a global minimum that would not otherwise exist. Designs that incorporate short stretches of D-amino acids were also validated using Rosetta's fragment-based

ab initio algorithm; the amino acid sequences of designs, with all D-amino acids mutated to glycine, were provided as input, and we allowed Rosetta to generate about 30,000 predicted structures as output. Unlike the standard *ab initio* protocol, we did not use secondary structure predictions in fragment picking. Additionally, the length of small and large fragments was set to 4 and 6 amino acid residues, respectively, instead of the default 3 and 9; we found that this produced better sampling for peptides. After conformational sampling, the D-amino acid positions were changed to their original identities and rescored. A small modification to the *ab initio* algorithm permitted it to build a terminal peptide bond for the N–C cyclic designs during the full-atom refinement stages of the structure prediction. Designs that showed no sampling near the design conformation or for which the design conformation was not the unique, lowest-energy conformation were discarded.

Because fragment-based methods are poorly suited to the prediction of structures with large amounts of D-amino acid content, such as NC₂CH₂HR₂D₁, we developed a new, fragment-free algorithm to validate these topologies. This algorithm, which we call *simple_cycpep_predict*, uses the same GenKIC-based sampling approach used to build backbones for design, with additional steps of filtering solutions on the basis of disulfide geometry, optimizing sidechain rotamers and gradient-descent energy minimization. Because the search space is vast, even with the constraints imposed by the N–C cyclic geometry and the disulfide bond(s), we further biased the search by setting mainchain torsion values for residues in the middle of the helices to helical values (a Gaussian distribution centred on $\phi = -61^\circ$, $\psi = -41^\circ$ for the α_R helix and on $\phi = +61^\circ$, $\psi = +41^\circ$ for the α_L helix); this is analogous to the biased sampling obtained by fragment-based methods, in which sequences with high helix propensity are sampled primarily with helical fragments. As with *ab initio* validation, designs showing poor sampling near the design conformation or poor energy landscapes were discarded.

Molecular-dynamics-based computational validation. We carried out further molecular-dynamics-based validation of the designs for which the *ab initio* or *simple_cycpep_predict* algorithms predicted high-quality energy landscapes. Similarly to strategies described previously^{33,34}, we used multiple short and independent trajectories, starting with different initial velocities to analyse the conformational flexibility and kinetic stability of the designed peptides. Molecular dynamics simulations were performed in explicit solvent conditions using the AMBER12 package and Amber ff12sb force field³⁵. A rectangular water box with a 10-Å buffer of TIP3P water³⁶ in each direction from the peptide was used for simulations. Sodium and chloride counterions were added to neutralize the system. The solvated system was minimized in two steps: solvent was first minimized for 20,000 cycles while keeping restraints on the peptide, followed by minimization of the whole system for another 20,000 cycles. At the start of simulations, the system was slowly heated from 0 K to 300 K under constant volume with positional restraints on the peptide of 10 kcal mol⁻¹ Å⁻¹ for 0.1 ns. For each selected peptide, 50 independent simulations starting with different initial velocities were performed. Each simulation started with the energy-minimized designed model, and was carried out for approximately 3.5 ns. Periodic boundary conditions were used with a constant temperature of 300 K using the Langevin thermostat³⁷ and a pressure of 1 atm with isotropic molecule-based scaling. A cut-off of 10 Å was used for the Lennard-Jones potential and the Particle Mesh Ewald method³⁸ to calculate long-range electrostatic interactions. The SHAKE algorithm³⁹ was applied to all bonds involving H atoms and an integration step of 2 fs was used for the simulations with amber12 PMEMD in the NPT ensemble. At the conclusion of the simulations, all the trajectories were analysed using the Amber12 package and VMD⁴⁰. We looked for fluctuations in root-mean-square deviation (r.m.s.d.), and for the convergence (or lack thereof) to the designed structure among all the trajectories. The distribution of r.m.s.d. values at the end of all trajectories was also analysed, although the beginning two-thirds of each trajectory were discarded as a burn-in period. Molecular dynamics analyses for three designs of the same topology are shown in Extended Data Fig. 4.

Prediction of mutational tolerance. Because the designed peptides presented here are intended to be used as starting points for designing binders to targets of therapeutic interest, we sought to examine the extent to which the designs can tolerate mutations (such as those that must be introduced to create a binding surface). Owing to the computational expense of the mutational analysis, we focused on the NC₂CH₂HR₂D₁ design, mutating each position in sequence to each of alanine, arginine, aspartate and phenylalanine, and carrying out a full structure prediction simulation for each. These mutations covered each class of mutation (elimination of the sidechain, introduction of a positive or negative charge, introduction of a bulky aromatic sidechain or introduction of a small aliphatic sidechain). Mutations preserved chirality; that is, only D-amino acid to D-amino acid or L-amino acid to L-amino acid mutations were considered. Simulation runs were carried out on the Argonne Leadership Computing Facility's Blue Gene/Q supercomputer ('Mira') using a version of the Rosetta *simple_cycpep_*

predict algorithm parallelized using the Message Passing Interface (MPI). The 127 prediction runs (each for a different mutation) each required approximately 20,000 CPU hours, and each produced about 25,000 sampled, closed conformations with good disulfide geometry. For each mutation considered, 50 trajectories were also carried out in which the mainchain was perturbed slightly and relaxed. The resulting collection of samples (from structure prediction and relaxation) was then used to calculate a goodness-of-energy-funnel metric, termed P_{near} :

$$P_{\text{near}} = \frac{\sum_{i=1}^N \exp\left(-\frac{\text{r.m.s.d.}_i^2}{\lambda^2}\right) \exp\left(-\frac{E_i}{k_B T}\right)}{\sum_{j=1}^N \exp\left(-\frac{E_j}{k_B T}\right)}$$

The value of P_{near} ranges from 0 (a poor funnel with low-energy alternative conformations or poor sampling close to the design conformation) to 1 (a funnel with a unique low-energy conformation very close to the design conformation). N is the number of samples, and E_i and r.m.s.d._{*i*} represent the Rosetta score and r.m.s.d. from the design structure of the *i*th sample, respectively. The parameter λ controls how close a state must be to the design if it is to be considered native-like; this was set to 1 Å. Similarly, the parameter $k_B T$ (where k_B is the Boltzmann constant and T is absolute temperature) governs the extent to which the shallowness or depth of the folding funnel affects the score; this was assigned a value of 1 Rosetta energy unit. The P_{near} metric provided a basis for comparison for the mutations considered.

Code availability. All the methods described here were implemented in the Rosetta software suite (<http://www.rosettacommons.org>). Rosetta software is freely available to academic and non-commercial users. Commercial licenses for the suite are available via the University of Washington Technology Transfer Office. Design protocols were implemented using the RosettaScripts interface available within the Rosetta software suite. Input files and command-line arguments for each step in our peptide design pipeline are available in Supplementary Information.

Protein purification of genetically encodable disulfide-rich peptides. Genes of designed disulfide-rich peptides were cloned into the vector pCDB180 (which we have made available via Addgene) using Gibson Assembly⁴¹. Protein expression from *E. coli* was carried out using a large N-terminal fusion domain consisting of the native *E. coli* protein OsmY to direct periplasmic and extracellular localization⁴², a deca-histidine tag for protein purification, and the SUMO protein Smt3 from *Saccharomyces cerevisiae* to chaperone folding and provide a mechanism for scarless cleavage of the fusion from the designed protein⁴³. Designed proteins were expressed from BL21*(DE3) *E. coli* (Invitrogen), and expression cultures were grown overnight with incubation at 37°C and shaking at 225 r.p.m. Following expression via Studier autoinduction⁴⁴, a periplasmic extract⁴⁵ was prepared by washing cells with 20% sucrose, 30 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 1 mg ml⁻¹ lysozyme. Protein was purified from the bacterial-conditioned medium and/or the periplasmic extract by immobilized metal-affinity chromatography (IMAC). During screening, fusion protein was purified from the bacterial-conditioned medium of 50 ml cultures, which typically yielded 9 ± 4 mg of protein (before removal of the fusion protein). Protein expression from mammalian cells was carried out using the Daedalus¹² system, as previously described. With both purification systems, purified fusion proteins were cleaved by site-specific proteases, SUMO protease for *E. coli* and TEV protease for Daedalus, followed by a secondary IMAC step. The final designs were purified to homogeneity by reverse-phase high-performance liquid chromatography on an Agilent 1260 HPLC equipped with a C-18 Zorbax SB-C18 4.6 mm × 150 mm column. Solvent A (water + 0.1% TFA) and solvent B (acetonitrile + 0.1% TFA) were applied using a gradient of 0%–45% solvent B ramping linearly at a rate of 1% per minute.

Synthesis and purification of non-canonical peptides. Linear and cyclic peptides were synthesized as previously described⁴⁶. Briefly, peptides were synthesized using automated solid-phase peptide synthesis with the Fmoc (9-fluorenylmethyloxycarbonyl) strategy. Cyclic reduced peptides were obtained after cleavage of the sidechain-protected peptides from the resin, ligation of both termini and the cleavage of sidechain protecting groups. Linear reduced peptides were collected by simultaneously cleaving the sidechain-protecting groups and the resin from the peptides. All linear or cyclic reduced peptides were oxidized at room temperature in a buffer containing 0.1 M NH₄HCO₃, in which the peptide concentration was 0.25 mg ml⁻¹. After 48 h, the mixture was acidified with trifluoroacetic acid, loaded onto a semi-preparative column and purified by RP-HPLC.

Mass spectrometry. Intact samples for each genetically encodable peptide were diluted in loading buffer with 0.1% formic acid and analysed on a Thermo Scientific Orbitrap Fusion Tribrid Mass Spectrometer via data-dependent acquisition. Liquid chromatography consisted of a 60-min gradient across a 15-cm column (internal diameter of 75 μm) packed with C₁₈ resin with a 3-cm kasil frit trap (internal

diameter of 150 μm) packed with C_{12} resin. For disulfide connectivity analysis, peptides were digested with sequencing-grade modified trypsin (Promega) at a 1:50 enzyme-to-substrate concentration for 1 h at 37 °C and then desalted via mixed-mode cationic exchange (MCX). Peptide samples were dried under vacuum and resuspended in 0.1% formic acid. Digested samples were analysed using data-dependent acquisition and targeted methods.

Thermal and chemical denaturation experiments. Circular dichroism (CD) wavelength and temperature scans were recorded on an AVIV model 420 or Jasco J-1500 CD spectrometer. For thermal denaturation, peptide samples were prepared at 0.07–0.2 mg ml^{-1} final concentration in 10 mM sodium phosphate buffer (pH 7.0). Wavelength scans from 195 nm to 260 nm were recorded at 25 °C, 55 °C, 95 °C and again after cooling back to 25 °C. For chemical denaturation experiments, samples for each peptide were prepared in the presence of 0–6 M GdnHCl concentrations. The concentration of GdnHCl was measured by refractometry⁴⁷. Peptide samples were also prepared in the presence of 2.5 mM TCEP (TCEP was pre-equilibrated to pH 7.0 before addition) and incubated for 3 h. Peptide concentrations were the same across all samples. Wavelength scans from 190 nm to 260 nm were recorded for each sample in a 0.1-cm cuvette.

NMR analysis and structure determination of genetically encodable disulfide-rich peptides. Agilent NMR spectrometers operating at ^1H resonance frequencies between 500 MHz and 750 MHz equipped with $^1\text{H}/^{15}\text{N}$, ^{13}C probes were used to acquire NMR data for gEHE_06, gEEHE_02, gEEH_04 and gHHH_06. The peptides were all uniformly ^{15}N -labelled. Peptide gEEH_04 was also about 10% labelled with ^{13}C . The peptides were dissolved in 50 mM sodium chloride, 20 mM sodium acetate, pH 4.8 (gEHE_06 and gEEHE_02) or 50 mM sodium phosphate, 4 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid, 0.02% sodium azide, pH 6.0 (gEEH_04 and gHHH_06). Final peptide concentrations ranged from 0.5 to 1.5 mM. The ^1H , ^{13}C and ^{15}N chemical shifts of the backbone and sidechain resonances were assigned by analysis of 2D [^{15}N , ^1H] HSQC, [^{13}C , ^1H] HSQC (aliphatic and aromatic), [^1H , ^1H] TOCSY and [^1H , ^1H] NOESY spectra, and 3D ^{15}N -resolved [^1H , ^1H] TOCSY, ^{15}N -resolved [^1H , ^1H] NOESY, HNCA, HNCO and HNHA spectra acquired at 20 °C (for gEHE_06 and gEEHE_02) or 25 °C (gEEH_04 and gHHH_06). Mixing times of 90 ms (gEHE_06 and gEEHE_02) and 200 ms (gEEH_04 and gHHH_06) were used for 2D and 3D NOESY, respectively. Slowly exchanging amides were identified for gEHE_06 and gEEHE_02 by lyophilizing a ^{15}N -labelled protein, re-dissolving in D_2O , and collecting a 2D [^{15}N , ^1H] HSQC spectrum about 10 min after re-dissolving the protein. The resulting D_2O sample was subsequently used to collect additional 2D [^1H , ^1H] TOCSY and [^1H , ^1H] NOESY data. Stereospecific assignments for the Val and Leu methyl groups were obtained for gEEH_04 for the 10% fractionally ^{13}C -labelled sample^{48,49}. Because it was not economical to prepare uniformly ^{13}C -labelled peptides by autoinduction, established triple-resonance NMR backbone assignment protocols could not be used. Instead, the carbon resonances were assigned by analysing the 2D [^1H , ^1H] TOCSY spectra along with [^{13}C , ^1H] HSQC spectra (collected at natural ^{13}C abundance for gHHH_06, gEHE_06 and gEEHE_02). For gEEH_04, which was 10% fractionally ^{13}C -labelled, the assignments were complemented with HNCA spectra. NMR data were processed using the Felix2007 (MSI) and PROSA (v6.4) programs and were analysed using the programs Sparky (v3.115), XEASY or CARA. Proton chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), whereas ^{13}C and ^{15}N chemical shifts were referenced indirectly via gyromagnetic ratios. Chemical shifts, NOESY peak lists and time-domain NMR data were deposited in the BioMagResBank (for accession numbers see Supplementary Table 2-1).

Isotropic overall rotational correlation times of 1.3–1.6 ns were inferred from averaged backbone ^{15}N spin relaxation times (<http://www.nmr2.buffalo.edu/negs-wiki>), indicating that all peptides are monomeric in solution. The ^1H , ^{13}C and ^{15}N chemical shift assignments and NOESY peak lists were used for iterative structure calculations using the program CYANA (v2.1 and v3.97). Chemical shifts were used to derive dihedral ϕ and ψ angle constraints using the program TALOS+⁵⁰ for residues located in well-defined regular secondary structure elements. For the final structure calculation, H-bond restraints¹³ were also introduced for gEHE_06 and gEEHE_02, for slowly exchanging amide protons. The resulting ensemble of 20 CYANA conformers was refined by restrained molecular dynamics in an ‘explicit water bath’ using the program CNS (v1.3)⁵¹. Structural quality was assessed using the online Protein Structure Validation Suite (PSVS, v1.5)⁵². The structural statistics are summarized in Supplementary Table 2-1. The coordinates for the 20 conformers representing the solution structures were deposited in the PDB (for accession numbers see Supplementary Table 2-1).

NMR analysis and structure determination of non-canonical peptides. Each non-canonical peptide (1 mg) was dissolved in 500 μl of 10% D_2O /90% H_2O or 100% D_2O (about pH 4). NMR spectra were recorded at 298 K on a Bruker Avance-600 spectrometer. Two-dimensional NMR experiments included TOCSY with an 80-s MLEV-17 spin lock, NOESY (mixing time of 200 ms), ECOSY and

natural-abundance ^{13}C and ^{15}N HSQC. Solvent suppression was achieved using excitation sculpting. Spectra were processed using Topspin 2.1 then analysed using CcpNmr Analysis⁵³. Chemical shifts were referenced to internal DSS.

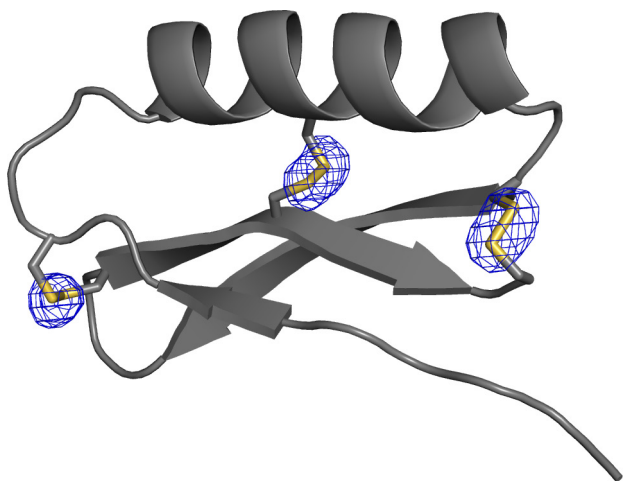
Initial structures were generated using CYANA and were based on distance restraints derived from NOESY spectra recorded in both 10% and 100% D_2O . The following restraints were also included: disulfide bonds; hydrogen bonds as indicated by slow D_2O exchange and sensitivity of amide proton chemical shift to temperature; χ_1 restraints from ECOSY and NOESY data; and backbone ϕ and ψ dihedral angles generated using the program TALOS-N⁵⁴. The final set of structures was generated within CNS⁵⁵ using torsion angle dynamics, refinement and energy minimization in explicit solvent, and protocols as developed for the RECOORD database⁵⁶. Final structures were assessed for stereochemical quality using MolProbity⁵⁷.

X-ray crystallography. The gEHEE_06 peptide was purified by size-exclusion chromatography on an AKTA Pure using a GE HiLoad 16/600 Superdex 75-pg column, concentrated to 50 mg ml^{-1} , and crystallized by vapour diffusion over well solutions of 100 mM citrate (pH 3.5) and 25% PEG3350. Selected crystals were transferred to a cryo-solution of 100 mM citrate (pH 3.5), 20% PEG3350 and 15% glycerol. Diffraction data were collected on a Rigaku Micromax-007HF with a Saturn944+ CCD detector, and integrated and scaled with HKL-2000. Initial phases were determined by molecular replacement using Phaser⁵⁸ as implemented in the CCP4 software suite with coordinates derived from a Rosetta model for the scaffold. Molecular replacement found two molecules per asymmetric unit. This solution was iteratively refined with the program Refmac followed by model building with COOT, yielding crystallographic R values of $R_{\text{cryst}} = 39.9\%$ and $R_{\text{free}} = 42.5\%$. On the basis of the Matthews’ coefficient, the crystals should have contained three molecules per asymmetric unit to have a reasonable solvent content of 45%. At this point, positive electron density appeared that enabled manual positioning of a third molecule in the asymmetric unit and improvement of the R values to $R_{\text{cryst}} = 32.0\%$ and $R_{\text{free}} = 34.9\%$. The model was further improved by including solvent molecules and TLS refinement. The quality of the final model was assessed using ProCheck and Molprobity (overall score: 100th percentile). The final model has been deposited in the PDB with accession code 5JG9. Crystallographic statistics are reported in Supplementary Table 2-2.

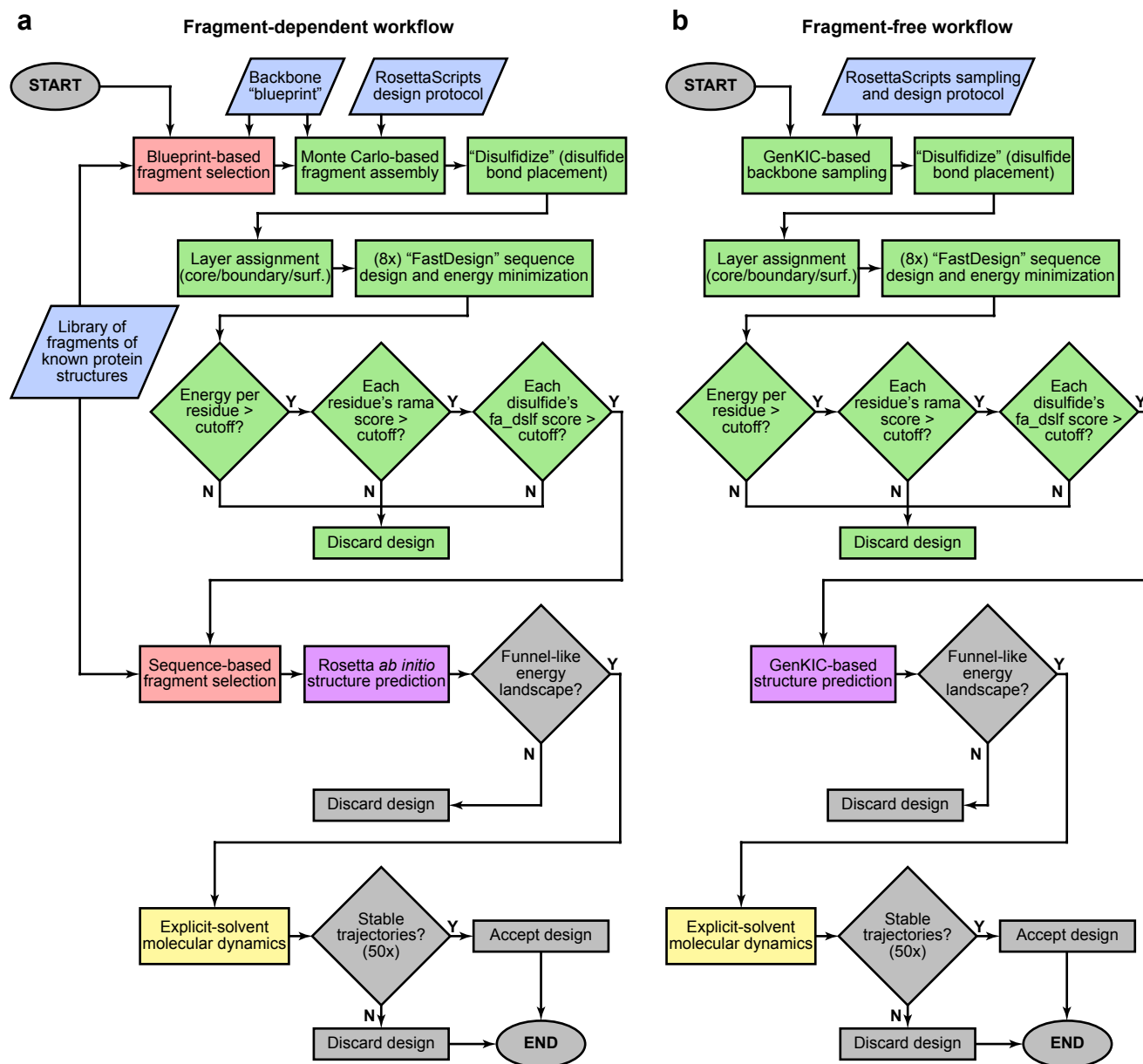
Surface redesign. In an attempt to reduce solubility and enhance crystallization, we redesigned solvent-exposed residues of designs representing each major topological category (mixed α/β , all β -sheet and all α -helical). Two resurfaced variants were selected for each design, bearing between one and two solvent-exposed tyrosine residues. We then expressed and purified these resurfaced designs using Daedalus, all but one of which expressed in a soluble manner and exhibited a redox-sensitive migration time by reverse-phase HPLC. We were able to obtain diffracting protein crystals for only redesign gEEHE_2.1_02_0008, which diffracted to 2.90-Å resolution (Supplementary Table 2-2). However, Matthews calculations predicted non-crystallographic symmetry with approximately 19 copies in the asymmetric unit, and attempts to phase the crystal by molecular replacement were unsuccessful, as were attempts to reproduce the crystal outside of the initial screen.

27. Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6**, e24109 (2011).
28. Lee, J., Lee, D., Park, H., Coutsiaris, E. A. & Seok, C. Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins* **78**, 3428–3436 (2010).
29. Harrison, P. M. & Sternberg, M. J. Analysis and classification of disulphide connectivity in proteins. The entropic effect of cross-linkage. *J. Mol. Biol.* **244**, 448–463 (1994).
30. Rodriguez-Granillo, A., Annavarapu, S., Zhang, L., Koder, R. L. & Nanda, V. Computational design of thermostabilizing d-amino acid substitutions. *J. Am. Chem. Soc.* **133**, 18750–18759 (2011).
31. O’Meara, M. J. *et al.* Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **11**, 609–622 (2015).
32. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution *de novo* structure prediction for small proteins. *Science* **309**, 1868–1871 (2005).
33. Caves, L. S., Evanseck, J. D. & Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **7**, 649–666 (1998).
34. Wijma, H. J. *et al.* Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, 49–58 (2014).
35. Case, D. A. *et al.* AMBER 12 <http://ambermd.org/doc12/Amber12.pdf> (Univ. California, 2012).
36. Jorgensen, W. L. & Corky, J. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. *J. Comput. Chem.* **19**, 1179–1186 (1998).
37. Loncharich, R. J., Brooks, B. R. & Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanine-N'-methylamide. *Biopolymers* **32**, 523–535 (1992).

38. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
39. Ryckaert, J.-P., Giovanni, C. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n -alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
40. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
41. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
42. Kotzsch, A. *et al.* A secretory system for bacterial production of high-profile protein targets. *Protein Sci.* **20**, 597–609 (2011).
43. Marblestone, J. G. *et al.* Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* **15**, 182–189 (2006).
44. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
45. Neu, H. C. & Heppel, L. A. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *J. Biol. Chem.* **240**, 3685–3692 (1965).
46. Cheneval, O. *et al.* Fmoc-based synthesis of disulfide-rich cyclic peptides. *J. Org. Chem.* **79**, 5538–5544 (2014).
47. Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280 (1986).
48. Neri, D. *et al.* Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. *Biochemistry* **28**, 7510–7516 (1989).
49. Herve du Penhoat, C. *et al.* The NMR solution structure of the 30S ribosomal protein S27e encoded in gene *RS27_ARCFU* of *Archaeoglobus fulgidis* reveals a novel protein fold. *Protein Sci.* **13**, 1407–1416 (2004).
50. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
51. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Alexandre, M. J. & Michael, N. Refinement of protein structures in explicit solvent. *Proteins Struct. Funct. Bioinf.* **50**, 496–506 (2003).
52. Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778–795 (2007).
53. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Bioinf.* **59**, 687–696 (2005).
54. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
55. Brünger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protocols* **2**, 2728–2733 (2007).
56. Nederveen, A. J. *et al.* RECOORD: a recalculated coordinate database of 500 proteins from the PDB using restraints from the BioMagResBank. *Proteins Struct. Funct. Bioinf.* **59**, 662–672 (2005).
57. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
58. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).



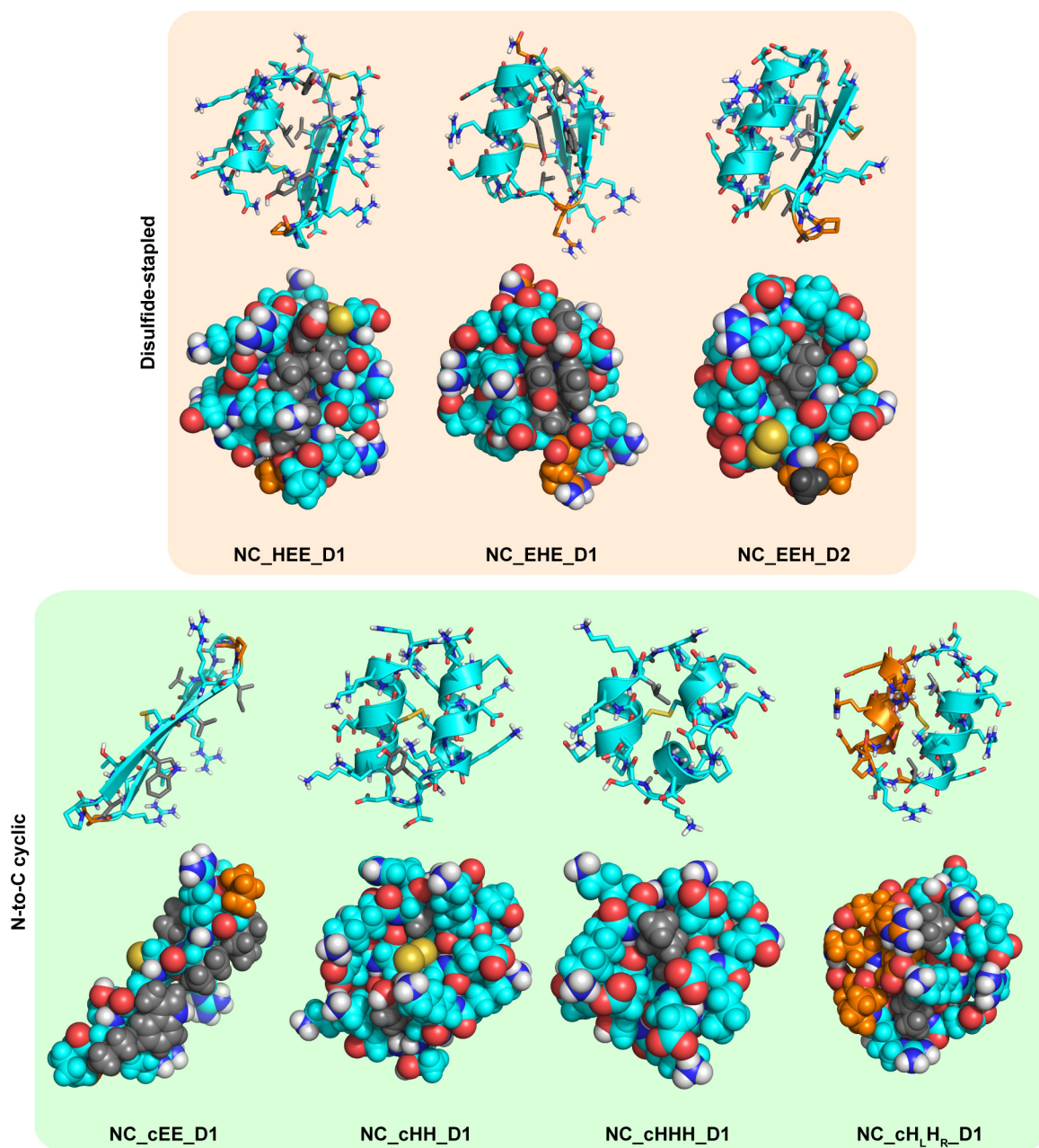
Extended Data Figure 1 | Disulfide bonds are well defined by X-ray crystallography. An $F_o - F_c$ omit-map is shown in blue, contoured at 4σ , for design gEHEE_06. Disulfide sulfur atoms were removed, and the omit-map was calculated following real-space refinement. The gEHEE_06 structure is shown in grey as a cartoon representation. Disulfide bonds are shown here as sticks, with sulfur atoms in yellow and carbon atoms in grey.



Extended Data Figure 2 | Flowchart of pipelines for designing non-canonical cyclic peptides. Inputs are shown in blue, RosettaScripts-automated parts of the pipeline are in green, parts carried out by Rosetta standalone applications are pink (the fragment picker application) and purple (the various structure prediction applications), parts performed with MD software are yellow, and manual steps are grey. **a**, Fragment-dependent design workflow. Final computational validation was carried out using MD simulations and fragment-based Rosetta *ab initio* structure

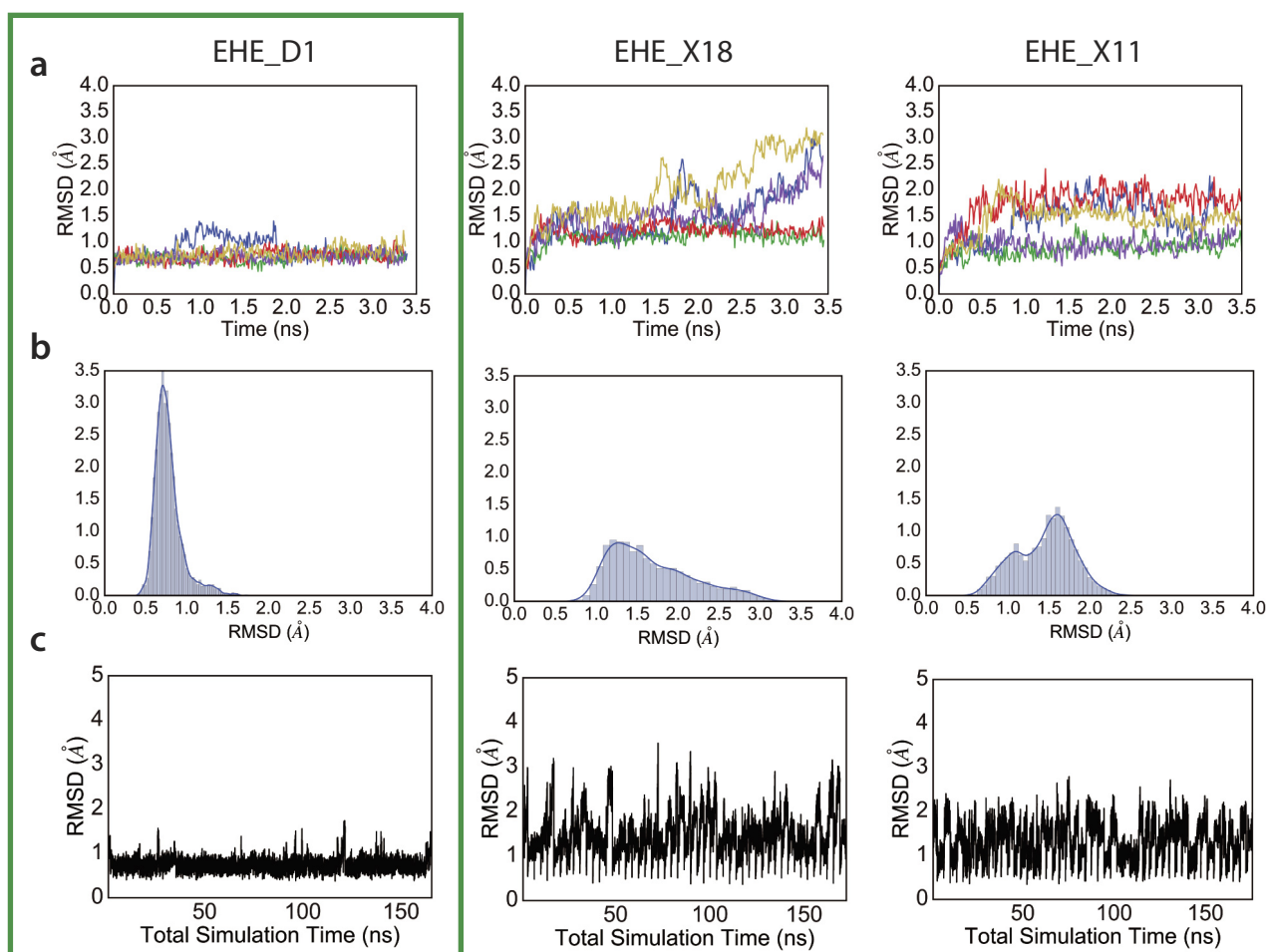
prediction. For peptides containing isolated D-amino acids, these residues were mutated to glycine for Rosetta *ab initio* structure prediction.

b, Fragment-free design workflow using GenKIC. This approach permits design of non-canonical topologies like the mixed $H_L H_R$ topology, which occurs in no known natural protein. The GenKIC-based structure prediction algorithm is described in Extended Data Fig. 7 and in Supplementary Information.



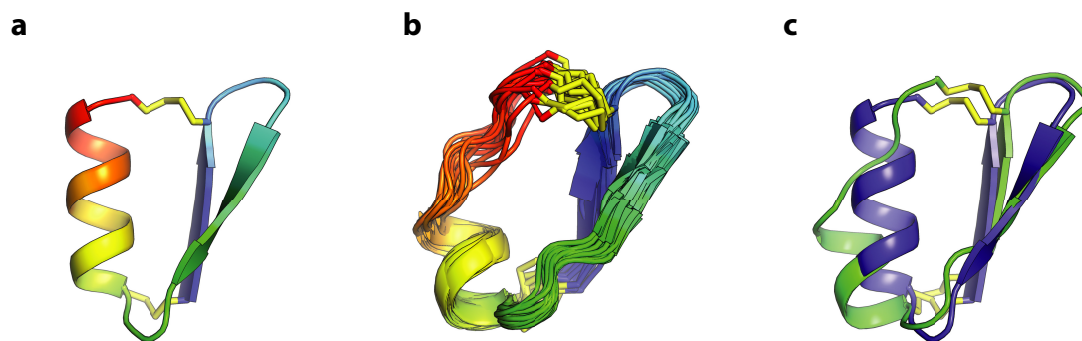
Extended Data Figure 3 | Sidechain placement in non-canonical peptide designs chosen for experimental characterization. Designs are shown as cartoon and stick representations (top row in each box) and as van der Waals spheres showing sidechain packing (bottom row in each box). L-amino acid residues are shown in cyan, and D-amino acid residues are

coloured orange. Sidechains of D- or L-variants of alanine, phenylalanine, isoleucine, leucine, valine, tryptophan and tyrosine are coloured grey to aid visualization of hydrophobic packing interactions. Top box, disulfide-stapled non-canonical peptide designs; bottom box, N-to-C cyclic non-canonical peptide designs.

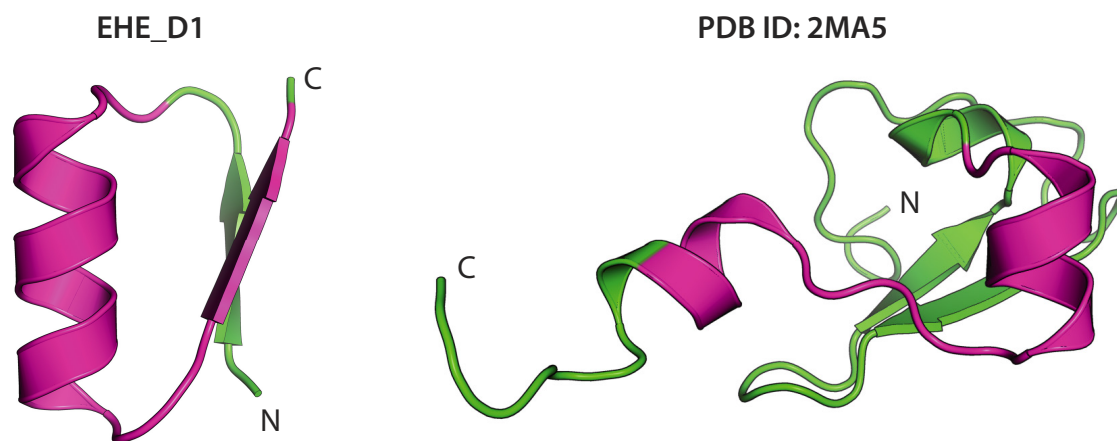


Extended Data Figure 4 | Molecular dynamics screening of designed peptides. Fifty independent molecular dynamics (MD) simulations in explicit solvent conditions, all starting from the designed peptide, were used for discriminating good, kinetically stable (for example, EHE_D1) designs from non-optimal designs of the same topology (for example, EHE_X18 and EHE_X11). **a**, Five representative trajectories from MD simulation runs. Designs that showed good convergence and smaller

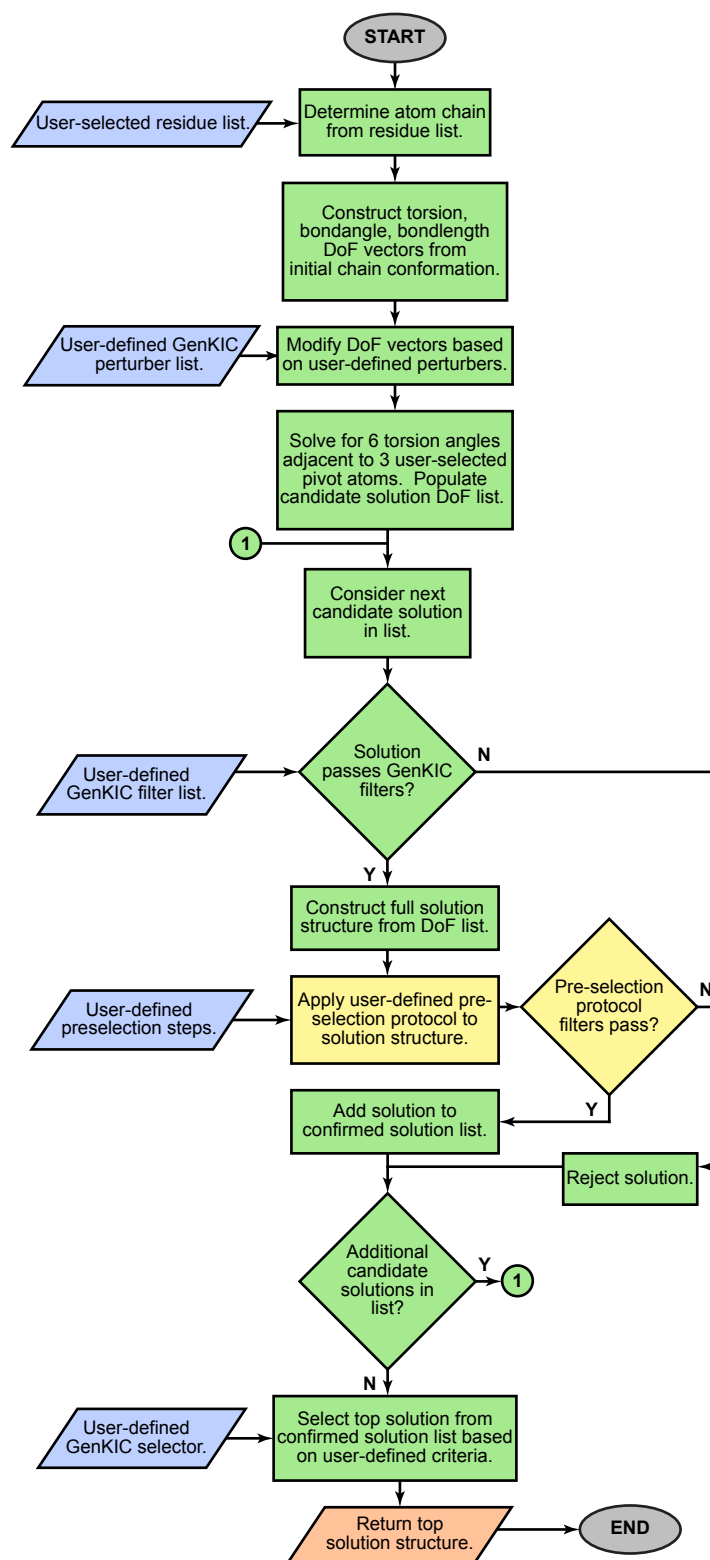
fluctuations were selected for further experimental characterization. **b**, r.m.s.d. distribution from all 50 trajectories. Blue line indicates the Gaussian kernel density estimate for the data. Only the last one-third of the trajectory was used for this analysis. Designs with narrower distributions were picked for further testing. **c**, Concatenated trajectory of all 50 independent runs show lower fluctuations for the more optimal designs.



Extended Data Figure 5 | Structural characterization of NC_EEH_D1. The NMR structure of NC_EEH_D1 does not match the designed topology. **a**, Rosetta-designed model for NC_EEH_D1. **b**, Ensemble of conformers representing the NMR solution structure. **c**, Superposition of the designed model (blue) with a representative NMR conformer (green).

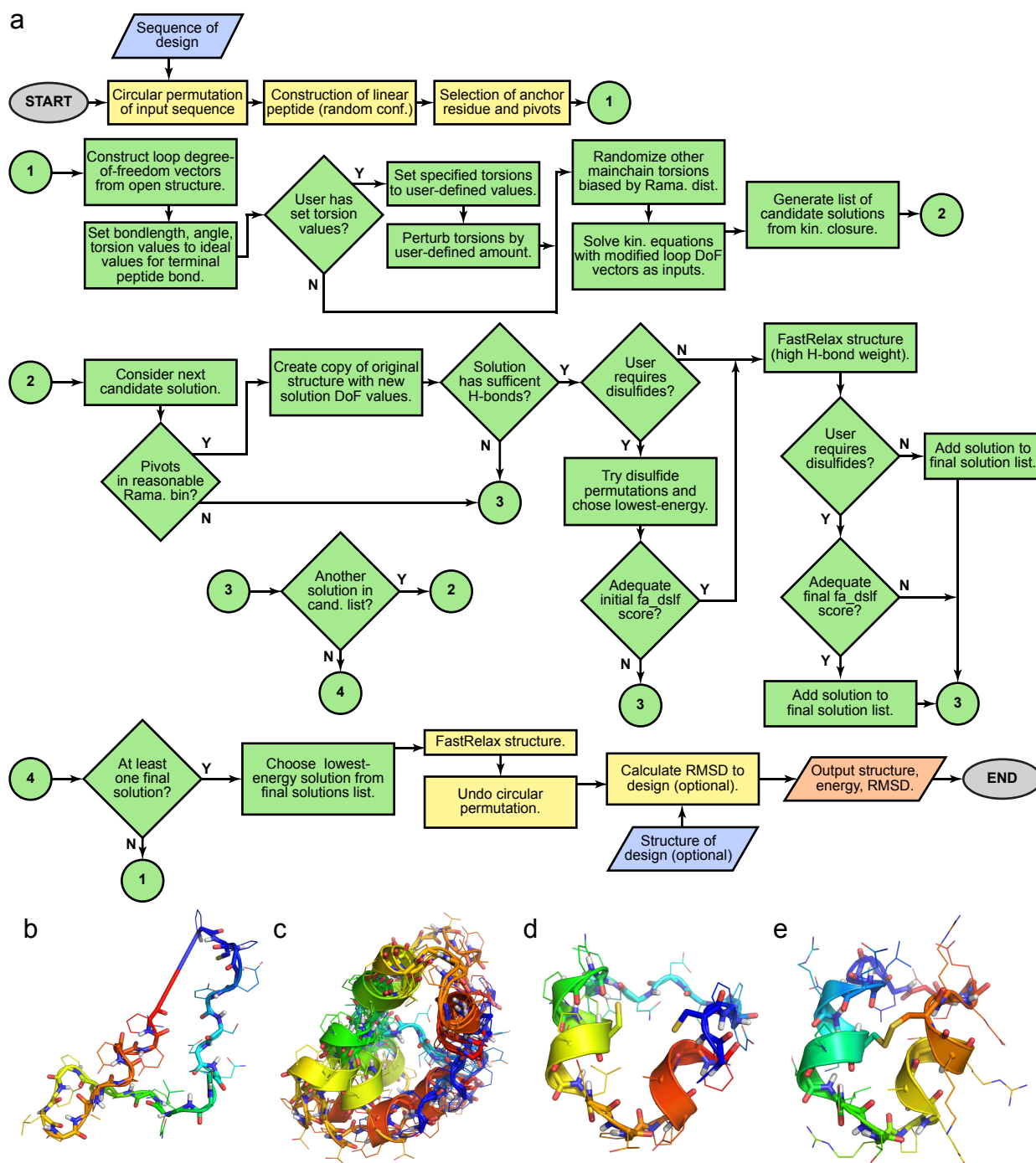


Extended Data Figure 6 | Structural mapping of sequence-aligned region between NC_EHE_D1 and 2MA5. Design NC_EHE_D1 and PDB entry 2MA5 show weak but significant (e -value, 2×10^{-4}) sequence alignment, which is highlighted in purple. The aligned region folds into very different structures in the different contexts of peptide and protein.



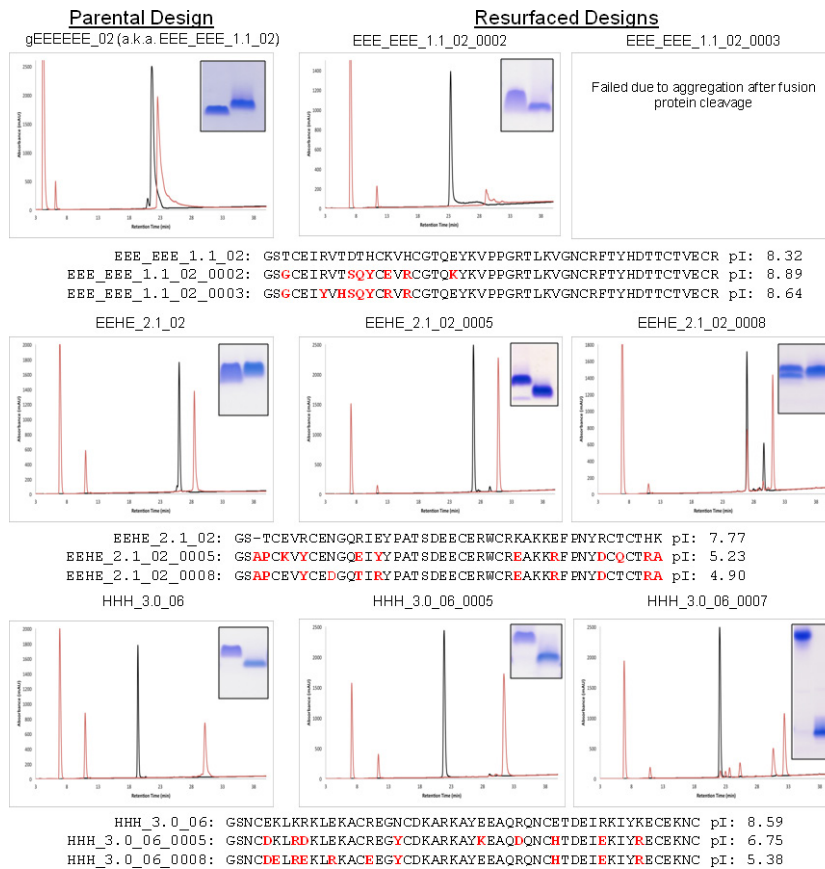
Extended Data Figure 7 | Generalized kinematic closure (GenKIC) algorithm flowchart. GenKIC allows sampling of closed conformations of arbitrary chains of atoms, passing through canonical or non-canonical backbone or sidechain linkages. Bond length, bond angle and torsional degrees of freedom in the chain can be fixed, perturbed from a starting value by small amounts, set to user-defined values, or sampled randomly. The algorithm then solves for six torsion angles adjacent to three user-defined pivot atoms in order to enforce closure of the loop. The many solutions from the closure are then filtered internally, and each can be

subjected to arbitrary user-defined Rosetta protocols and filtration in order to prune the solution list further. A single solution is selected from those passing filters by a user-defined selection criterion. This flowchart shows the steps in a single invocation of the algorithm; for sampling, a user may specify that the algorithm be applied any number of times. User inputs are shown in blue, steps carried out by the GenKIC algorithm itself are in green, steps carried out by Rosetta code external to the GenKIC algorithm are shown in yellow, and outputs are shown in salmon.



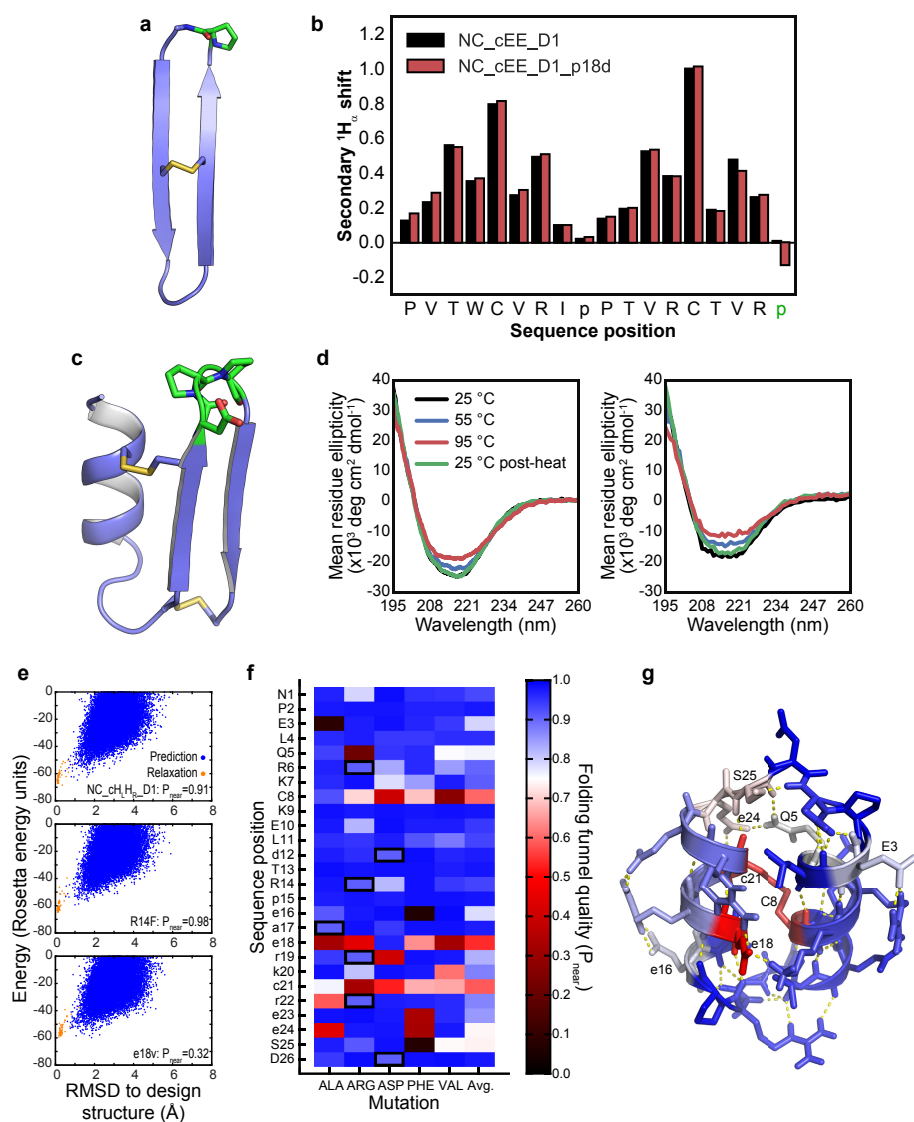
Extended Data Figure 8 | A new fragment-free structure prediction algorithm. **a**, Flowchart of the steps required to generate a single sampled conformation. In typical usage, this process would be repeated tens of thousands of times to produce many samples. Inputs (the peptide sequence and an optional PDB file for the design structure) are shown in blue, and outputs (the sampled structure, its energy, and its r.m.s.d. from the design structure) are shown in salmon. Steps performed by the GenKIC algorithm are shaded green, and setup and completion steps performed by the *simple_cycpep_predict* application are shown in yellow. Further details of this algorithm are discussed in Supplementary Information.

b, The initial, random peptide conformation with bad terminal peptide bond geometry. **c**, Ensemble of closed conformations found for a single closure attempt. In this example, residue 7 (cyan) is the fixed anchor residue. Certain regions of the peptide have been set to left- or right-handed helical conformations before solving closure equations. **d**, A single closed solution with relative cysteine sidechain orientations that pass the initial, low-stringency filter for disulfide (fa_dslf) conformational energy. **e**, The resulting structure, following sidechain repacking, energy minimization, and cyclic de-permutation.



Extended Data Figure 9 | Mutational tolerance of selected genetically-encodable designs. Left column, RP-HPLC traces for the parental designs; middle and right, same for the resurfaced designs where applicable. Traces for proteins run under oxidizing conditions are shown as black lines, while traces for proteins run following reduction with 10 mM DTT are shown as

red lines. Insets, gels highlighting the SDS-PAGE mobility of each purified protein under oxidizing (left band) and reducing conditions (right band). Under each row of panels are shown sequence alignments with the mutated positions highlighted in red, along with theoretical isoelectric points as calculated by ProtParam.



Extended Data Figure 10 | Mutational tolerance of selected NC designs.

a, b, Mutational tolerance of the D-proline, L-proline loop of design NC_cEE_D1 (green in **a**), assessed by secondary $^1\text{H}_\alpha$ chemical shift (p.p.m.) for the design sequence (black bars in **b**) and the p18d loop mutation (red bars). Eliminating this key proline residue does not result in loss of β -strand signal. **c, d**, Mutational tolerance of loop region of design NC_HEE_D1 (green in **c**), as assessed by CD spectroscopy for the design sequence (left plot in **d**) and for the D19T, p20q, P21D triple mutant (right plot in **d**). Both proline residues may be mutated without loss of secondary structure or major change in the thermal stability. **e–g**, Computationally predicted mutational tolerance of design NC_HLHR_D1, across the entire sequence. Each position was successively mutated *in silico* to D- or L-alanine, arginine, aspartate, phenylalanine, or valine (preserving the position's chirality), and full folding simulations were carried out with the Rosetta *simple_cycpep_predict* application. Folding funnel quality was evaluated using the P_{near} metric described

in Methods. **e**, Representative plots of energy versus r.m.s.d. from the design structure, plotted for the design sequence (top), for the non-disruptive R14F mutation (middle), and for the e18v mutation (bottom). Results from GenKIC-based structure prediction runs are shown in blue, and relaxation runs, in orange. Note that the bottom case shows many sampled states far from the design state with energy equal to or less than the design state energy. **f**, Mutational tolerance by position (vertical axis) and mutation (horizontal axis). Blue rectangles represent well-tolerated mutations, and red to black rectangles represent disruptive mutations, based on P_{near} evaluation of the folding funnel. Black borders indicate the design sequence. **g**, Mutational tolerance mapped onto the NC_HLHR_D1 structure, with colours as in **f**. Most positions tolerate mutation well, with only the disulfide bridge (C8–c21) and the salt bridges formed by e18 being highly sensitive. The hydrogen bond networks formed by residues Q5, e24 and s25 show some moderate sensitivity to mutation, as do residues E3 and e16.

Genome evolution in the allotetraploid frog *Xenopus laevis*

Adam M. Session^{1,2*}, Yoshinobu Uno^{3*}, Taejoon Kwon^{4,5*}, Jarrod A. Chapman², Atsushi Toyoda⁶, Shuji Takahashi⁷, Akimasa Fukui⁸, Akira Hikosaka⁹, Atsushi Suzuki⁷, Mariko Kondo¹⁰, Simon J. van Heeringen¹¹, Ian Quigley¹², Sven Heinz¹³, Hajime Ogino¹⁴, Haruki Ochi¹⁵, Uffe Hellsten², Jessica B. Lyons¹, Oleg Simakov¹⁶, Nicholas Putnam¹⁷, Jonathan Stites¹⁷, Yoko Kuroki¹⁸, Toshiaki Tanaka¹⁹, Tatsuo Michiue²⁰, Minoru Watanabe²¹, Ozren Bogdanovic²², Ryan Lister²², Georgios Georgiou¹¹, Sarita S. Paranjpe¹¹, Ila van Kruijsbergen¹¹, Shengquiang Shu², Joseph Carlson², Tsutomu Kinoshita²³, Yuko Ohta²⁴, Shuuji Mawaribuchi²⁵, Jerry Jenkins^{2,26}, Jane Grimwood^{2,26}, Jeremy Schmutz^{2,26}, Therese Mitros¹, Sahar V. Mozaffari²⁷, Yutaka Suzuki²⁸, Yoshikazu Haramoto²⁹, Takamasa S. Yamamoto³⁰, Chiyo Takagi³⁰, Rebecca Heald³¹, Kelly Miller³¹, Christian Haudenschild^{32†}, Jacob Kitzman³³, Takuya Nakayama³⁴, Yumi Izutsu³⁵, Jacques Robert³⁶, Joshua Fortriede³⁷, Kevin Burns³⁷, Vaneet Lotay³⁸, Kamran Karimi³⁸, Yuuri Yasuoka³⁹, Darwin S. Dichmann¹, Martin F. Flajnik²⁴, Douglas W. Houston⁴⁰, Jay Shendure³³, Louis DuPasquier⁴¹, Peter D. Vize³⁸, Aaron M. Zorn³⁷, Michihiko Ito⁴², Edward M. Marcotte⁴, John B. Wallingford⁴, Yuzuru Ito²⁹, Makoto Asashima²⁹, Naoto Ueno^{30,43}, Yoichi Matsuda³, Gert Jan C. Veenstra¹¹, Asao Fujiyama^{6,44,45}, Richard M. Harland¹, Masanori Taira⁴⁶ & Daniel S. Rokhsar^{1,2,16}

To explore the origins and consequences of tetraploidy in the African clawed frog, we sequenced the *Xenopus laevis* genome and compared it to the related diploid *X. tropicalis* genome. We characterize the allotetraploid origin of *X. laevis* by partitioning its genome into two homoeologous subgenomes, marked by distinct families of ‘fossil’ transposable elements. On the basis of the activity of these elements and the age of hundreds of unitary pseudogenes, we estimate that the two diploid progenitor species diverged around 34 million years ago (Ma) and combined to form an allotetraploid around 17–18 Ma. More than 56% of all genes were retained in two homoeologous copies. Protein function, gene expression, and the amount of conserved flanking sequence all correlate with retention rates. The subgenomes have evolved asymmetrically, with one chromosome set more often preserving the ancestral state and the other experiencing more gene loss, deletion, rearrangement, and reduced gene expression.

Ancient polyploidization events have shaped diverse eukaryotic genomes¹, including two rounds of whole-genome duplication at the base of the vertebrate radiation². While polyploidy is rare in amniotes,

presumably owing to constraints on sex chromosome dosage^{3,4}, it is common in fish⁵, amphibians^{6,7}, and plants⁸. Polyploidy provides raw material for evolutionary diversification because gene duplicates

¹University of California, Berkeley, Department of Molecular and Cell Biology and Center for Integrative Genomics, Life Sciences Addition #3200, Berkeley, California 94720-3200, USA. ²US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. ³Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. ⁴Department of Molecular Biosciences, Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, USA. ⁵Department of Biomedical Engineering, School of Life Sciences, Ulsan National Institute of Science and Technology, Ulsan 689-798, Republic of Korea. ⁶Center for Information Biology, and Advanced Genomics Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. ⁷Amphibian Research Center, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. ⁸Laboratory of Tissue and Polymer Sciences, Faculty of Advanced Life Science, Hokkaido University, N10W8, Kita-ku, Sapporo 060-0810, Japan. ⁹Division of Human Sciences, Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8521, Japan. ¹⁰Misaki Marine Biological Station (MMBS), Graduate School of Science, The University of Tokyo, 1024 Koaji, Misaki, Miura, Kanagawa 238-0225, Japan. ¹¹Radboud University, Faculty of Science, Department of Molecular Developmental Biology, 259 RIMLS, M850/2.97, Geert Grooteplein 28, Nijmegen 6525 GA, the Netherlands. ¹²Salk Institute, Molecular Neurobiology Laboratory, La Jolla, San Diego, California 92037, USA. ¹³Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, San Diego, California 92037, USA. ¹⁴Department of Animal Bioscience, Nagahama Institute of Bio-Science and Technology, 1266 Tamura, Nagahama, Shiga 526-0829, Japan. ¹⁵Institute for Promotion of Medical Science Research, Yamagata University Faculty of Medicine, 2-2-2 Iida-Nishi, Yamagata, Yamagata 990-9585, Japan. ¹⁶Molecular Genetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan. ¹⁷Dovetail Genomics LLC, Santa Cruz, California 95060, USA. ¹⁸Department of Genome Medicine, National Research Institute for Child Health and Development, NCHD, 2-10-1, Okura, Setagaya-ku, Tokyo 157-8535, Japan. ¹⁹Department of Life Science and Technology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan. ²⁰Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1, Komaba, Meguro-ku, Tokyo 153-8902, Japan. ²¹Institute of Institution of Liberal Arts and Fundamental Education, Tokushima University, 1-1 Minamijosanjima-cho, Tokushima 770-8502, Japan. ²²Harry Perkins Institute of Medical Research and ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, Perth, Western Australia 6009, Australia. ²³Department of Life Science, Faculty of Science, Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan. ²⁴Department of Microbiology and Immunology, University of Maryland, 655 W Baltimore St, Baltimore, Maryland 21201, USA. ²⁵Kitasato Institute for Life Sciences, Kitasato University, 5-9-1 Shirokane Minato-ku, Tokyo 108-8641, Japan. ²⁶HudsonAlpha Institute of Biotechnology, Huntsville, Alabama 35806, USA. ²⁷Department of Human Genetics, University of Chicago, 920 E. 58th St, CLSC 431F, Chicago, Illinois 60637, USA. ²⁸Department of Computational Biology and Medical Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan. ²⁹Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology (AIST), Central 5, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8565, Japan. ³⁰Division of Morphogenesis, Department of Developmental Biology, National Institute for Basic Biology, 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan. ³¹University of California, Berkeley, Department of Molecular and Cell Biology, Life Sciences Addition #3200, Berkeley California 94720-3200, USA. ³²Illumina Inc., 25861 Industrial Blvd, Hayward, California 94545, USA. ³³Department of Genome Sciences, University of Washington, Foege Building S-250, Box 355065, 3720 15th Ave NE, Seattle Washington 98195-5065, USA. ³⁴Department of Biology, University of Virginia, Charlottesville, Virginia 22904, USA. ³⁵Department of Biology, Faculty of Science, Niigata University, 8050, Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan. ³⁶Department of Microbiology & Immunology, University of Rochester Medical Center, Rochester, New York 14642, USA. ³⁷Division of Developmental Biology, Cincinnati Children's Research Foundation, Cincinnati, Ohio 45229-3039, USA. ³⁸Department of Biological Sciences, University of Calgary, Alberta T2N 1N4, Canada. ³⁹Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan. ⁴⁰The University of Iowa, Department of Biology, 257 Biology Building, Iowa City, Iowa 52242-1324, USA. ⁴¹Department of Zoology and Evolutionary Biology, University of Basel, Basel CH-4051, Switzerland. ⁴²Department of Biological Sciences, School of Science, Kitasato University, 1-15-1 Minamiku, Sagami-hara, Kanagawa 252-0373, Japan. ⁴³Department of Basic Biology, SOKENDAI (The Graduate University for Advanced Studies), 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan. ⁴⁴Principles of Informatics, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. ⁴⁵Department of Genetics, SOKENDAI (The Graduate University for Advanced Studies), 1111 Yata, Mishima, Shizuoka 411-8540, Japan. ⁴⁶Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. †Present address: Personalis Inc., 1330 O'Brien Drive, Menlo Park, California 94025, USA. *These authors contributed equally to this work.

can support new functions and networks⁹. However, the component subgenomes of a polyploid must cooperate to mediate potential incompatibilities of dosage, regulatory controls, protein–protein interactions and transposable element activity.

The African clawed frog *X. laevis* is one of a polyploid series that ranges from diploid to dodecaploid, and is therefore ideal for studying the impact of genome duplication¹⁰, especially given its status as a model for cell and developmental biology¹¹. *X. laevis* has a chromosome number ($2n=36$) nearly double that of the Western clawed frog *Xenopus* (formerly *Silurana*) *tropicalis* ($2n=20$) and most other diploid frogs¹², and is proposed to be an allotetraploid that arose via the interspecific hybridization of diploid progenitors with $2n=18$, followed by subsequent genome doubling to restore meiotic pairing and disomic inheritance^{10,13} (see Supplementary Note 1 and Extended Data Fig. 1 for discussion of the *Xenopus* allotetraploidy hypothesis).

Here we provide evidence for the allotetraploid hypothesis by tracing the origins of the *X. laevis* genome from its extinct progenitor diploids. The two subgenomes are distinct and maintain separate recombinational identities. Despite sharing the same nucleus, we find that the subgenomes have evolved asymmetrically: one of the two subgenomes

has experienced more intrachromosomal rearrangement, gene loss by deletion and pseudogenization, and changes in levels of gene expression and in histone and DNA methylation. Superimposed on these global trends are local gene family expansions and the alteration of gene expression patterns.

Assembly, annotation and karyotype

We sequenced the genome of the *X. laevis* inbred 'J' strain by whole-genome shotgun methods in combination with long-insert clone-based end sequencing, (Supplementary Note 2) and organized the assembled sequences into chromosomes using fluorescence *in situ* hybridization (FISH) of 798 bacterial artificial chromosome clones (BACs) and *in vivo* and *in vitro* chromatin conformation capture analysis (Supplementary Note 3 and Methods). These complementary methods produced a high-quality chromosome-scale draft that includes all previously known *X. laevis* genes and assigns >91% of the assembled sequence (and 90% of the predicted protein-coding genes) to a chromosomal location.

We annotated 45,099 protein-coding genes and 342 microRNAs using RNA sequencing (RNA-seq) from 14 developmental

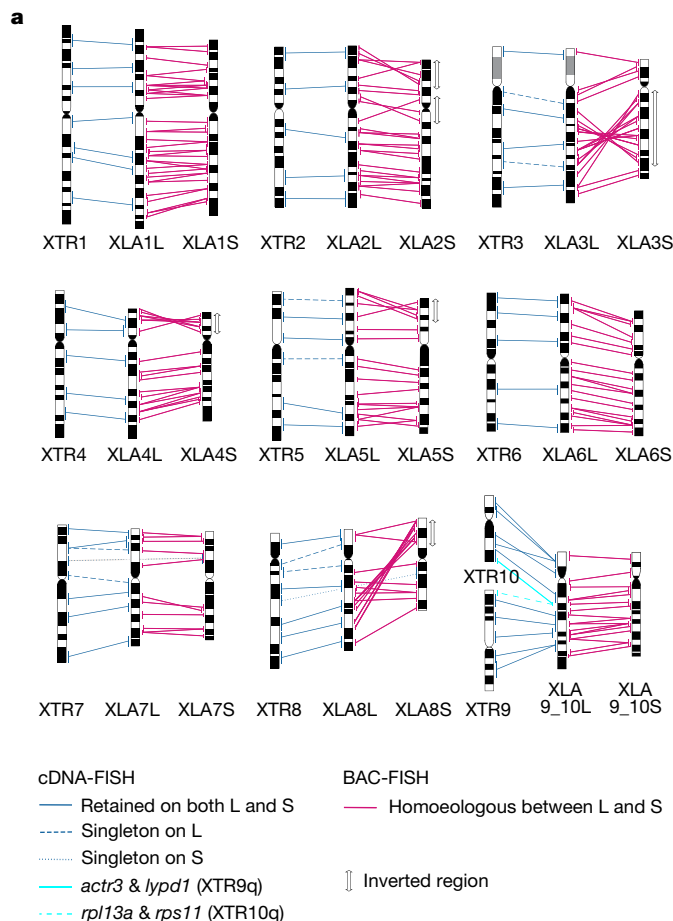
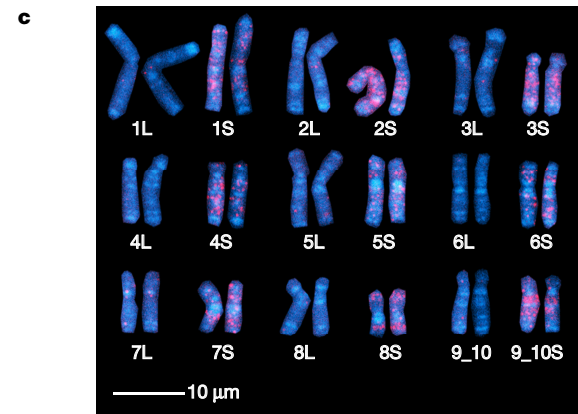
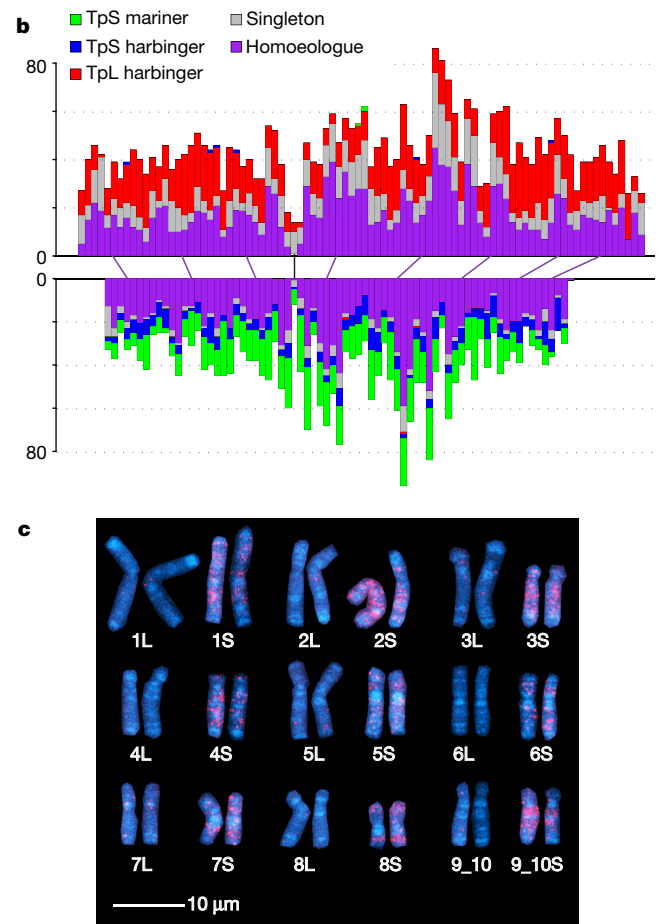


Figure 1 | Chromosome evolution in *Xenopus*. **a**, Comparative cytogenetic map of XLA (*Xenopus laevis*) and XTR (*Xenopus tropicalis*) chromosomes. Magenta lines show relationships of chromosomal locations of 198 homoeologous gene pairs between XLA.L and XLA.S chromosomes, identified by FISH mapping using BAC clones (Supplementary Table 1 and Supplementary Note 3.1). Blue lines show relationships of chromosomal locations of orthologous genes between XTR chromosomes and (i) both XLA.L and XLA.S chromosomes (solid line) (lines between XLA.L and XLA.S are omitted), (ii) only XLA.L (dashed), or (iii) only XLA.S (dotted), which were taken from our previous studies^{14,15}. Light blue lines indicate positional relationships of *actr3* and *lypd1* on XTR9q and *rpl13a* and *rps11* on XTR10q with those on XLA9_10LS chromosomes (Supplementary Note 6.2). Double-headed arrows on the right of XLA.S chromosomes



indicate the chromosomal regions in which inversions occurred. Ideograms of XTR and XLA chromosomes were taken from our previous reports^{15,16}. **b**, Distribution of homoeologous genes (purple), singletons (grey) and subgenome-specific repeats across XLA1L (top) and XLA1S (bottom). Xl-TpL_harb is red, Xl-TpS_harb is blue, and Xl-TpS_mar is green. Purple lines mark homoeologous genes present in both L and S chromosomes, the black line marks the approximate centromere location on each chromosome. The homoeologous gene pairs, from left to right: *rnf4*, *spcs3*, *intsl2*, *foxa1*, *sds*, *ap3s1*, *lifr*, *aqp7*. Each bin is 3 Mb in size, with 0.5 Mb overlap with the previous bin. **c**, Chromosomal localization of the Xl-TpS_mar sequence with fluorescence *in situ* hybridization. Hybridization signals were only observed on the S chromosomes. Scale bar, 10 μm.

stages (including the oocyte stage) and 14 adult tissues and organs (Supplementary Note 4), analysis of histone marks associated with transcription, and homology with *X. tropicalis* and other tetrapods (Supplementary Note 5 and Methods). Of the *X. laevis* protein-coding genes, 24,419 can be placed in 2:1 or 1:1 correspondence with 15,613 *X. tropicalis* genes, defining 8,806 homoeologous pairs of *X. laevis* genes with *X. tropicalis* orthologues and 6,807 single copy orthologues. The remaining genes are members of larger gene families (such as olfactory receptor genes) whose *X. tropicalis* orthology is more complex.

The *X. laevis* karyotype (Fig. 1a) reveals nine pairs of homoeologous chromosomes^{1,14,15}. Each of the first eight pairs is co-orthologous to and named for a corresponding *X. tropicalis* chromosome, appending an L and S for the longer and shorter homoeologues, respectively¹⁶. XLA2L is the Z/W sex chromosome¹⁷, for which we determined a W-specific sequence in the q-subtelomeric region that includes the sex-determining gene *dmw*¹⁷ and a corresponding Z-specific haplotype. The homoeologous XLA2Sq, by contrast, has no such locus, and neither does XTR2 (Extended Data Fig. 2a and Supplementary Note 6). The ninth pair of homoeologues is a q-q fusion of proto-chromosomes homologous to XTR9 and XTR10, which probably occurred before allotetraploidization (Extended Data Fig. 2b–d and Supplementary Note 6). The S chromosomes are, on average, 13.2% shorter karyotypically¹⁶ and 17.3% shorter in assembled sequence than their L counterparts. The single nucleotide polymorphism rate in *X. laevis* is approximately 0.4%, far less than the approximately 6% divergence between homoeologous genes (Extended Data Fig. 1c and Supplementary Note 8.8).

Subgenomes and timing of allotetraploidization

We reasoned that dispersed relicts of transposable elements specific to each progenitor would mark the descendent subgenomes in an allotetraploid (Fig. 2c and Extended Data Fig. 1). Three classes of DNA transposon relicts appeared almost exclusively on either the L or S chromosomes (Supplementary Note 7). Xl-TpL_harb and Xl-TpS_harb are subfamilies of miniature inverted-repeat transposable elements (MITE) of the PIF/harbinger superfamily^{18,19} whose relicts were almost completely restricted to L or S chromosomes, respectively (Fig. 1b and Extended Data Fig. 3a). Similarly, sequence relicts of the Tc1/mariner superfamily member Xl-TpS_mar (closely related to the fish MMTS subfamily²⁰) were found almost exclusively on the S chromosomes (Fig. 1b), as confirmed by FISH analysis using Xl-TpS_mar as a probe (Fig. 1c and Supplementary Note 7.4; see Supplementary Note 7.3 for details on the rare elements that map to the opposite subgenome).

The L and S chromosome sets therefore represent the descendants of two distinct diploid progenitors, confirming the allotetraploid hypothesis despite the absence of extant progenitor species. Analysis of synonymous divergence of protein-coding genes suggests that the L and S subgenomes diverged from each other around 34 Ma (T_2) and from *X. tropicalis* around 48 Ma (T_1) (Fig. 2a), consistent with prior gene-by-gene estimates from transcriptomes^{21–24} (Supplementary Note 8, Extended Data Fig. 4 and Methods). L- and S-specific transposable elements were active around 18–34 Ma, indicating that the two progenitors were independently evolving diploids during that period (Fig. 2a, Supplementary Note 7.5 and Extended Data Fig. 3). More recent transposon activity is more uniformly distributed across the L and S chromosomes (not shown). Finally, consistent with a common origin for tetraploid *Xenopus* species, we can clearly identify orthologues of L and S genes in whole-genome sequences of a related allotetraploid frog, *X. borealis*, and estimate the *X. laevis*–*X. borealis* divergence to be around 17 Ma (T_3). These considerations constrain the allotetraploid event to around 17–18 Ma (T^*). This timing is consistent with other estimates of the radiation of tetraploid *Xenopus* species, which are presumed to emerge from the bottleneck of a shared allotetraploid founder population^{23,24}.

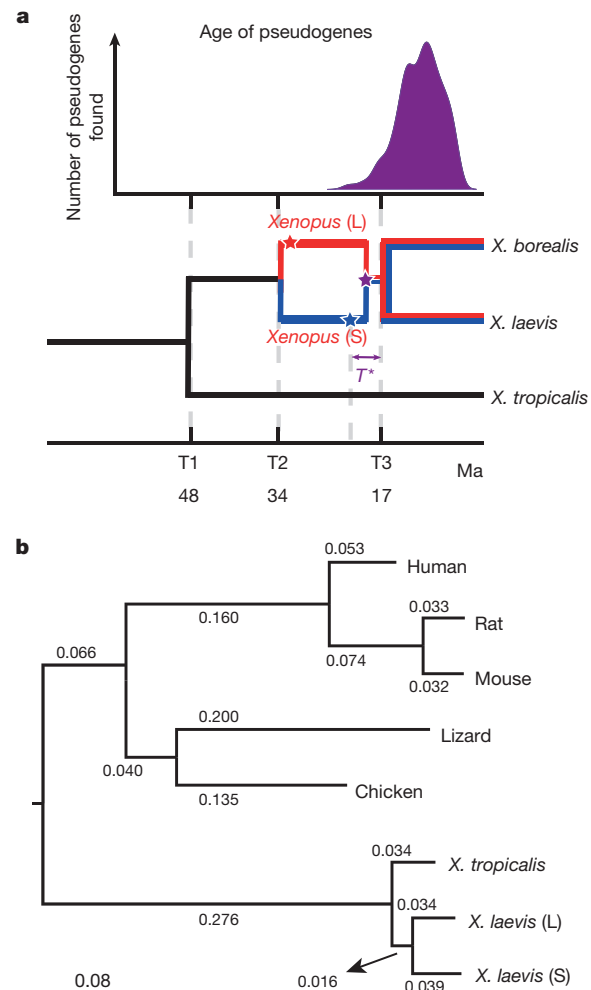


Figure 2 | Molecular evolution and allotetraploidy. **a**, The distribution of pseudogene ages, as described in Supplementary Note 9 (top). Phylogenetic tree illustrating the different epochs in *Xenopus* (bottom), with times based on protein-coding gene phylogeny of pipids, including *Xenopus*, *Pipa carvalhoi*, *Hymenochirus boettgeri* and *Rana pipiens* (only *Xenopus* depicted). We date the speciation of *X. tropicalis* and the *X. laevis* ancestor at 48 Ma, the L and S polyploid progenitors at 34 Ma and the divergence of the polyploid *Xenopus* radiation at 17 Ma. Using these times as calibration points, we estimate bursts of transposon activity at 18 Ma (mariner, blue star) and 33–34 Ma (harbinger, red star). The purple star is the time of hybridization, around 17–18 Ma. **b**, Phylogenetic tree based on protein-coding genes of tetrapods, rooted by elephant shark (not shown). Alignments were done by MACSE (multiple alignment of coding sequences accounting for frameshifts and stop codons) and the maximum-likelihood tree was built by PhyML. Branch length scale shown at the bottom for 0.08 substitutions per site. The difference in branch length between *Xenopus laevis*-L and *Xenopus laevis*-S is similar to that seen between mouse and rat. Both subgenomes of *X. laevis* have longer branch lengths than *X. tropicalis*.

Karyotype stability

With the exception of the chromosome 9–10 fusion, *X. laevis* and *X. tropicalis* chromosomes have maintained conserved synteny since their divergence around 48 Ma (Fig. 1a, b). The absence of inter-chromosomal rearrangements is consistent with the relative stability of amphibian and avian karyotypes compared to those of mammals²⁵, which typically show dozens of inter-chromosome rearrangements²⁶. It also contrasts with many plant polyploids, which can show considerable inter-subgenome rearrangement²⁷. The distribution of L- and S-specific repeats along entire chromosomes implies the absence of crossover recombination between homoeologues since allotetraploidization, presumably because the two progenitors were sufficiently diverged to

Table 1 | Summary of retention of different genomic elements, in comparison to the diploid *X. tropicalis* genome

Sequence element	XTR	XLA.L	XLA.S	Retention
Protein coding genes	15,613	13,781	10,241	56.4%
Genomic DNA (Mb)	1,227	1,222	1,010	N/A
microRNAs (miRNAs)	180	166	168	86.7%
Pan vertebrate conserved noncoding elements	550	542	536	96.6%
H3K4me3 peaks	7,473	6,927	5,833	70.6%
p300 peaks	4,321	3,457	2,702	42.5%
Cactus	1,294,342	1,026,204	888,899	49.0%
MitoCarta	917	717	501	46.0%
GermPlasm	15	15	6	40.0%

More detailed information is available in Supplementary Tables 2 and 3. XTR, *X. tropicalis*; XLA.L, *X. laevis* L; XLA.S, *X. laevis* S.

avoid meiotic pairing between homoeologous chromosomes, though we cannot rule out very limited localized inter-homoeologue exchanges (Supplementary Note 7).

The extensive collinearity between homologous *X. laevis* L and *X. tropicalis* chromosomes (Fig. 1a) implies that they represent the ancestral chromosome organization. In contrast, the S subgenome shows extensive intra-chromosomal rearrangements, evident in the large inversions of XLA2S, XLA3S, XLA4S, XLA5S and XLA8S, as well as shorter rearrangements (Fig. 1a). The S subgenome has also experienced more deletions. For example, the 45S pre-ribosomal RNA gene cluster is found on *X. laevis* XLA3Lp, but its homoeologous locus on XLA3Sp is absent (Extended Data Fig. 5a). Extensive small-scale deletions (Extended Data Fig. 5b) reduce the length of S chromosomes relative to their L and *X. tropicalis* counterparts (see below).

Response of subgenomes to allotetraploidy

Redundant functional elements in a polyploid are expected to rapidly revert to single copies through the fixation of disabling mutations and/or loss²⁸ unless prevented by neofunctionalization⁸, subfunctionalization²⁶, or selection for gene dosage²⁹. Differential gene loss between homoeologous chromosomes is sometimes referred to as 'genome fractionation'^{30–32} (Supplementary Note 1). At least 56.4% of the protein-coding genes duplicated by allotetraploidization have been retained in the *X. laevis* genome (Supplementary Note 10; 60.2% if genes on unassigned short scaffolds are included). Previous studies that relied on cDNA²¹ and expressed sequence tag (EST) surveys^{22,33,34} observed far lower rates of retention, probably owing to sampling biases from gene expression (Supplementary Note 8.2).

Even higher retention rates were found for homoeologous microRNAs (156 out of 180, 86.7%), similar to the salmonid-specific duplication⁵, and both primary copies are expressed for intergenic homoeologous microRNAs (Supplementary Note 8.6 and Extended Data Fig. 5e). Pan-vertebrate putatively *cis*-regulatory conserved non-coding elements (CNEs)³⁵ were also highly retained (541 out of 550, 98.4%; Supplementary Note 8.7 and Table 1). CNEs conserved between *X. laevis* and *X. tropicalis*, however, were retained at a significantly lower rate (49%, $P \leq 1 \times 10^{-50}$; Table 1 and Supplemental Table 3). Longer genes (by genomic span, exon number or coding length) were more likely to be retained (Wilcoxon signed-rank test, $P \leq 10^{-5}$; Supplementary Note 10.5 and Extended Data Fig. 5 h–j), broadly consistent with the idea that longer genes have more independently mutable functions and are therefore more susceptible to subfunctionalization and subsequent retention³⁶.

Genes have been lost asymmetrically between the two subgenomes of *X. laevis*. Similar results have been reported for some plant polyploids³⁰ but not in rainbow trout⁵. For *X. laevis* protein-coding genes with clear 1:1 or 2:1 orthologues in *X. tropicalis*, we found that significantly more genes were lost from the S subgenome (31.5%) than from the L subgenome (8.3%; χ^2 test $P = 2.23 \times 10^{-50}$; Supplementary Table 2), with

the same trend for other types of functional elements, such as histone H3 lysine 4 trimethylation (H3K4me3)-enriched promoters and p300-bound enhancers (Table 1). Across most of the genome, genes appeared to be lost independently of their neighbours, as runs of gene losses were nearly geometrically distributed (Fig. 3a, right). We did observe some large block deletions (for example, several olfactory clusters (Extended Data Fig. 5b) and a few unusually long blocks of functionally unrelated genes that were retained in two copies without loss (Fig. 3a, left)).

Many lost genes were simply deleted, as demonstrated by significantly shorter distances between conserved flanking genes in the other subgenome and in *X. tropicalis*. Both the size and number of deletions were greater on the S subgenome (Extended Data Fig. 5c). We identified 985 'unitary' (that is, non-retrotransposed) pseudogenes out of 1,531 loci examined in detail. This 64% detection rate was similar between subgenomes in *X. laevis* and comparable to that reported in trout⁵. Based on the accumulation of non-synonymous mutations³⁷ we estimated that most of these pseudogenes escaped evolutionary constraint around 15 Ma (Fig. 2a and Extended Data Fig. 6), consistent with the onset of extensive redundancy in the allotetraploid, although the precision of our pseudogene age estimates is low (Supplementary Note 9). Most pseudogenes showed no evidence of expression, but of 769 pseudogenes longer than 100 bp, 133 (17.2%) showed residual expression (Extended Data Fig. 6). Conversely, among homoeologous gene pairs, we found 760 for which one member had little to no expression across our 28 sampled conditions. Although these retained some gene structure (start and stop codon, no frame shifts, good splice signals), they showed increased rates of amino acid change and appeared to be under relaxed selection (Extended Data Fig. 5f). We called these nominally dying genes 'thanagenes' (Supplementary Note 12.5). Reduced expression may be due to mutated *cis*-regulatory elements, as exemplified by the *six6* gene pair (Fig. 4e, Extended Data Fig. 8 g–i and Supplementary Note 13.1).

Although tetraploidy created two 'copies' of nearly every gene, additional gene copies were continually produced by tandem duplication (Fig. 3d and Extended Data Fig. 7). The number of tandem clusters was greater in *X. tropicalis* than in the *X. laevis* L subgenome, which in turn was greater than in the S subgenome (Supplementary Note 11.1). Although tandem duplication was faster in *X. tropicalis* than in *X. laevis*, there was also a higher rate of loss. Since tandem duplications and deletions occur by unequal crossing over during meiosis, these differing rates were consistent with the shorter generation time of *X. tropicalis* (Extended Data Fig. 7 f, g). The mean time to loss of an old tandem duplicate is around 40 Ma in *X. laevis* (on either subgenome) compared to around 16 Ma in *X. tropicalis*. Homoeologous gene loss and tandem duplication can combine to yield complex histories for some gene families. We discuss how these families contribute to the literature on whole-genome duplication evolution in Supplementary Notes 10 and 13.

Functional patterns of gene retention and loss

We found preferential retention or loss of many functional categories (Fig. 4a, Extended Data Figs 4e, 9, 10 and Supplementary Note 13). DNA binding proteins, components of developmentally regulated signalling (TGF β , Wnt, Hedgehog and Hippo) and cell cycle regulation pathways were retained at a substantially higher rate (>90%) than average (Extended Data Fig. 10). Genes retained in multiple copies after the ancient vertebrate genome duplication were also more likely to be retained as homoeologues in *X. laevis* (Supplementary Note 10.4), similar to teleost and trout genome duplications⁵. We found nearly complete retention of 37 out of 38 duplicated genes in the four pairs of homoeologous Hox clusters, with a single pseudogene (Fig. 3c). High rates of homoeologue retention in most genes in these categories suggest that stoichiometrically controlled expression levels may be needed, or subfunctionalization of homoeologues may have occurred, either in their expression domain or in their target specificity.

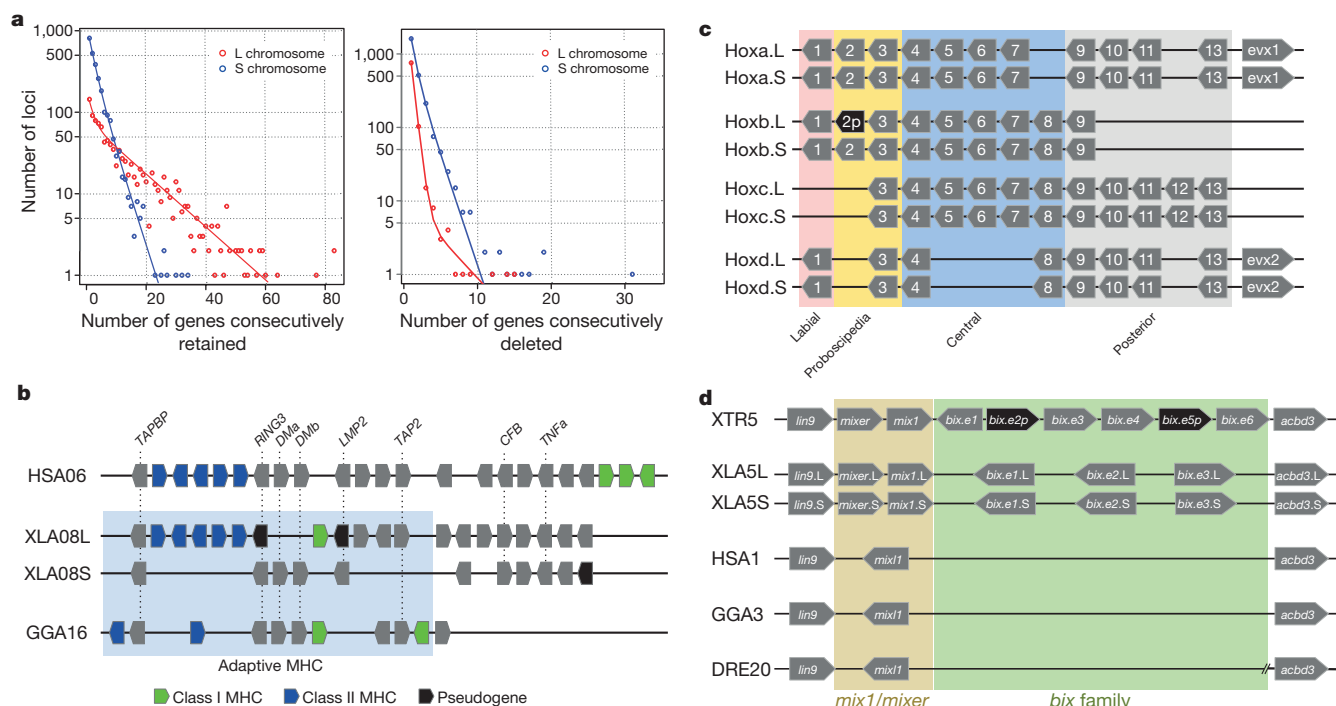


Figure 3 | Structural response to allotetraploidy. **a**, Distributions of consecutive retentions (left) and deletions (right) in the L (red) and S (blue) subgenomes. The distributions were fit using the equation $y = a \times (e^{bx}) + c \times (e^{dx})$. The y axis is shown on a log scale. Significant differences were seen between L and S subgenomes in both distributions (Student's t -test, retention, $P = 3.6 \times 10^{-22}$; deletion, $P = 4.5 \times 10^{-84}$). **b**, Evolutionary conservation of the *Xenopus* major histocompatibility complex (MHC) and differential MHC silencing on the two *X. laevis* subgenomes. Selected gene names shown above. The 'Adaptive MHC' encodes tightly-linked essential genes involved in antigen presentation to T cells; this group of genes is the primordial linkage group and has been preserved in most non-mammalian vertebrates, including *Xenopus*. Differential gene silencing is particularly pronounced, as four genes around the class I gene are functional on the S chromosome, but absent (*dma* and *dmb* (MHC-class II domain alpha and beta) or pseudogenes (*ring3*, really interesting new gene 3; *lmp2*, large multifunctional

Conversely, homoeologous genes in other functional categories have been lost at a higher rate, presumably because of a corresponding lack of selection for dosage. For example, genes involved in DNA repair were lost at a high rate (79%) (Supplementary Note 10.1). This is consistent with relaxed selection for repair in the immediate aftermath of allotetraploidy, when all genes were present in four copies per somatic cell⁵. Other metabolic categories were also prone to loss, presumably because single loci encoding enzymes were sufficient³⁸. Genomic regions with notable loss include the major histocompatibility complex genes on the S subgenome (Fig. 3b) and several olfactory receptor clusters (Extended Data Fig. 5b). We hypothesize that homoeologous genes may be functionally incompatible in these cases, leading to en bloc deletion in response to selection pressure. Specific case studies of duplicate gene retention and loss are detailed in Extended Data Figs 9, 10 and Supplementary Note 13.

Evolution of gene expression

Gene expression is also a predictor of retention, whereby more highly expressed genes are more likely to be retained (Extended Data Fig. 8b), similar to results seen in *Paramecium*^{39,40}. Developmentally regulated genes whose expression levels peak at the maternal zygotic transition (MZT) or during neural differentiation were retained at higher levels ($P < 0.01$), based on gene expression networks constructed from developmental and adult tissue expression (Methods, Fig. 4a (right), Extended Data Fig. 10e and Supplementary Note 12.3). We speculate

that the exceptional retention of developmentally regulated genes is due to selection for stoichiometric dosage of these factors, and in some cases higher expression in the physically larger allotetraploid cells and embryos relative to those of diploid frogs, although a propensity³⁶ of these genes for sub- or neofunctionalization could also have contributed. In the adult, genes whose expression peaks in the brain and eye were also retained at higher levels (Fig. 4b). In *X. laevis*, the expression of homoeologues was highly correlated (Extended Data Fig. 8a), showing that the overall expression of homoeologues diverged similarly to that of orthologues between *Xenopus* species⁴¹. Many homoeologous pairs, however, were differentially expressed throughout development or across adult tissues, either in a spatiotemporal pattern (a form of subfunctionalization³⁶; Supplementary Note 12.4 and Extended Data Fig. 8d–f) or in the same pattern but with differing expression levels. When homoeologous gene pairs were both expressed, the average L copy expression level was approximately 25% higher than that of the S copy consistently across adult tissues and after the MZT⁴² (Fig. 4b and Supplementary Note 12.2). Excess L expression, however, averaged only around 12% in oocyte and early pre-MZT stages, suggesting that the two subgenomes were more evenly expressed as maternal transcripts but developed an increased asymmetry after the MZT. Strikingly, we found 391 cases in which one homoeologue had no detectable maternal mRNA (oocytes, egg and stage 8; Fig. 4c, d and Extended Data Fig. 8c). Compared to similar transcript data from *X. tropicalis*, we found cases of an apparent

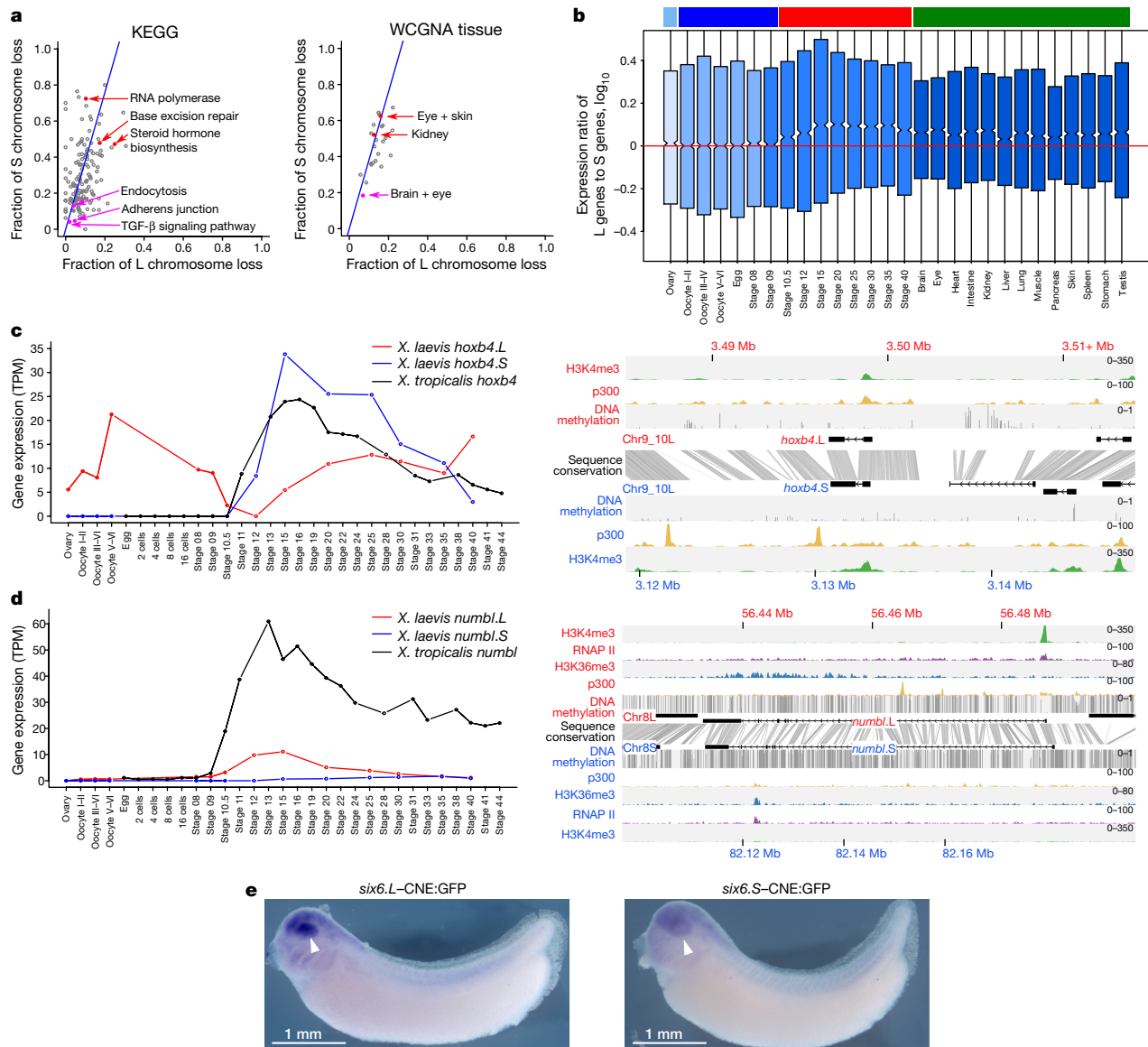


Figure 4 | Retention and functional differentiation. **a**, Comparison of L and S gene loss by KEGG categories (left) and tissue-weighted gene co-expression network analysis (WGCNA) categories (right) (Supplementary Note 10.1). Blue line denotes expected L or S loss based on genome-wide average (56.4%). Red points denote functional categories showing a high degree of loss. Magenta points denote functional categories showing a high degree of retention (χ^2 test, $P < 0.01$). **b**, Box plot of $\log_{10}(LTPM/STPM)$ for homoeologous gene pairs, zoomed in to show medians. Ovary and maternally controlled developmental time points (left, light blue and dark blue bars, respectively), zygotically controlled developmental time points and adult tissues (right, red and green bars, respectively). Red line, equal ratio $\log_{10}(1)$. On average, maternal datasets express the L gene of a homoeologous pair 12% more strongly than the S gene (median = 0%), whereas zygotically controlled developmental time points and adult tissues express the L gene of a homoeologous pair 25% more strongly than the S gene (median = 1.8%). The difference between the mean and medians is explained by many genes with large differences between homoeologues (Extended Data Fig. 8c). **c**, **d**, Developmental expression plot (left)

loss of expression ('maternal subfunctionalization', that is, *X. tropicalis* and one *X. laevis* gene was expressed, whereas the other *X. laevis* gene was silenced pre-MZT) in 238 genes (for example *numbl.S*). We also found gains of expression ('maternal neofunctionalization', that is, the *X. tropicalis* gene was not expressed maternally, but one *X. laevis* gene was expressed) in 153 genes (for example *hoxb4.L*). We did not see such

and epigenetic landscape (right) surrounding *hoxb4* (c) and *numbl* (d). Right, genomic profiles of H3K4me3 (green), p300 (yellow), RNA polymerase II (RNAP II; purple) and H3K36me3 (blue) ChIP-seq tracks, as well as DNA methylation levels determined by whole-genome bisulfite sequencing (grey). Gene annotation track shows *hoxb4* (c) and *numbl* (d) genes on L (top) and S. Grey denotes conservation between L and S genomic sequences. **d**, The small amount of expression seen in maternal *numbl* and *numbl.L* is consistent between replicates. Gene expression is measured in transcripts per million mapped reads (TPM). **e**, Representative embryos with GFP expression, as detected by *in situ* hybridization at stages 32–33, driven by *six6.L*-CNE or *six6.S*-CNE linked to a basal promoter-GFP cassette (*six6.L*-CNE:GFP and *six6.S*-CNE:GFP, respectively). Embryos were 4,250–4,450 μ m. Semi-quantitative image analysis revealed a substantial difference in average expression level; the expression driven by *six6.S*-CNE ($n = 27$) was 0.6-fold weaker than that by *six6.L*-CNE in the eye region ($n = 32$). Given eye-specific patterns of their endogenous expression, the *six6* genes probably have additional silencers for restricting enhancer activity of the CNEs in the eye.

a large divergence in other expression domains (Supplementary Note 12.2 and Extended Data Fig. 8c), suggesting a higher level of plasticity of maternal mRNA regulation between *X. laevis* homoeologues, similar to the trend seen between *Xenopus* species⁴¹.

Overall, thousands of homoeologue pairs have either divergent spatiotemporal patterns or similar patterns with differing expression

levels. Such homoeologue pairs differed in substitution rate and coding sequence length difference more than those that were similar in expression (Supplementary Note 12.4 and Extended Data Fig. 8d–f), a pattern that was also found in trout homoeologous pairs⁵. These expression differences can largely be attributed to changes in epigenetic regulation (Random Forest classification; ROC area under the curve 0.78), with changes in H3K4me3 and DNA methylation contributing the most explanatory power among our epigenetic variables (Supplementary Note 14). Detailed comparison of the two subgenomes will facilitate identification of specific sequences that control *cis*-regulatory differences between homoeologues.

Conclusion

The two subgenomes of *Xenopus laevis* have evolved asymmetrically, with the L subgenome more consistently resembling the ancestral condition and the S subgenome more disrupted by deletion and rearrangement. Asymmetric gene loss has been observed in allopolyploid plants³⁰ and yeast⁴³ at the segmental level, but it has not been shown directly that similarly fractionated segments derive from the same progenitor (Fig. 1c). Our results are consistent with the idea that optimized gene expression levels are an important force affecting gene retention following polyploidy^{39,40}. The asymmetry between the L and S subgenomes could have been the result of an intrinsic difference between their diploid progenitors. Alternately, the remodelling of the S genome could have been a response to the L–S merger itself, a ‘genomic shock’⁴⁴ resulting from the activation of transposable elements (Fig. 2a and Supplementary Note 8.5). The popularity of *Xenopus* as a model for the study of vertebrate development, cell biology and immunology is now extended to a model for vertebrate polyploidy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 December 2015; accepted 9 September 2016.

1. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009).
2. Holland, P. W., Garcia-Fernández, J., Williams, N. A. & Sidow, A. Gene duplications and the origins of vertebrate development. *Development Suppl.*, 125–133 (1994).
3. Muller, H. J. Why polyploidy is rarer in animals than in plants. *Am. Nat.* **59**, 346–353 (1925).
4. Orr, H. A. ‘Why polyploidy is rarer in animals than in plants’ revisited. *Am. Nat.* **136**, 759–770 (1990).
5. Berthelot, C. *et al.* The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* **5**, 3657 (2014).
6. Woods, I. G. *et al.* The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* **15**, 1307–1314 (2005).
7. Glasauer, S. M. K. & Neuhauss, S. C. F. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics* **289**, 1045–1060 (2014).
8. Otto, S. P. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
9. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
10. Kobel, H. R. & Du Pasquier, L. Genetics of polyploid *Xenopus*. *Trends Genet.* **2**, 310–315 (1986).
11. Harland, R. M. & Grainger, R. M. *Xenopus* research: metamorphosed by genetics and genomics. *Trends Genet.* **27**, 507–515 (2011).
12. Kuramoto, M. A list of chromosome numbers of anuran amphibians. *Bull. Fukuoka Univ. Educ.* **39**, 83–127 (1990).
13. Bisbee, C. A., Baker, M. A., Wilson, A. C., Haji-Azimi, I. & Fischberg, M. Albumin phylogeny for clawed frogs (*Xenopus*). *Science* **195**, 785–787 (1977).
14. Uno, Y., Nishida, C., Takagi, C., Ueno, N. & Matsuda, Y. Homeologous chromosomes of *Xenopus laevis* are highly conserved after whole-genome duplication. *Heredity* **111**, 430–436 (2013).
15. Uno, Y. *et al.* Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One* **7**, e33027 (2012).
16. Matsuda, Y. *et al.* A new nomenclature of *Xenopus laevis* chromosomes based on the phylogenetic relationship to *Silurana/Xenopus tropicalis*. *Cytogenet. Genome Res.* **145**, 187–191 (2015).
17. Yoshimoto, S. *et al.* A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl Acad. Sci. USA* **105**, 2469–2474 (2008).

18. Zhang, X. *et al.* P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc. Natl Acad. Sci. USA* **98**, 12572–12577 (2001).
19. Jurka, J. & Kapitonov, V. V. PIFs meet Tourists and Harbingers: a superfamily reunion. *Proc. Natl Acad. Sci. USA* **98**, 12315–12316 (2001).
20. Ahn, S. J., Kim, M.-S., Jang, J. H., Lim, S. U. & Lee, H. H. MMTS, a new subfamily of Tc1-like transposons. *Mol. Cells* **26**, 387–395 (2008).
21. Morin, R. D. *et al.* Sequencing and analysis of 10,967 full-length cDNA clones from *Xenopus laevis* and *Xenopus tropicalis* reveals post-tetraploidization transcriptome remodeling. *Genome Res.* **16**, 796–803 (2006).
22. Hellsten, U. *et al.* Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Biol.* **5**, 31 (2007).
23. Bewick, A. J., Chain, F. J. J., Heled, J. & Evans, B. J. The pipid root. *Syst. Biol.* **61**, 913–926 (2012).
24. Cannatella, D. *Xenopus* in space and time: fossils, node calibrations, tip-dating, and paleobiogeography. *Cytogenet. Genome Res.* **145**, 283–301 (2015).
25. Voss, S. R. *et al.* Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res.* **21**, 1306–1312 (2011).
26. Ferguson-Smith, M. A. & Trifonov, V. Mammalian karyotype evolution. *Nat. Rev. Genet.* **8**, 950–962 (2007).
27. Langham, R. J. *et al.* Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics* **166**, 935–945 (2004).
28. Haldane, J. B. S. The part played by recurrent mutation in evolution. *Am. Nat.* **67**, 5–19 (1933).
29. Birchler, J. A. & Veitia, R. A. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl Acad. Sci. USA* **109**, 14746–14753 (2012).
30. Schnable, J. C., Springer, N. M. & Freeling, M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl Acad. Sci. USA* **108**, 4069–4074 (2011).
31. Sankoff, D., Zheng, C. & Wang, B. A model for biased fractionation after whole genome duplication. *BMC Genomics* **13** (Suppl. 1), S8 (2012).
32. Garsmeur, O. *et al.* Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* **31**, 448–454 (2014).
33. Sémon, M. & Wolfe, K. H. Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proc. Natl Acad. Sci. USA* **105**, 8333–8338 (2008).
34. Chain, F. J. J., Dushoff, J. & Evans, B. J. The odds of duplicate gene persistence after polyploidization. *BMC Genomics* **12**, 599 (2011).
35. Lee, A. P., Kerk, S. Y., Tan, Y. Y., Brenner, S. & Venkatesh, B. Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. *Mol. Biol. Evol.* **28**, 1205–1215 (2011).
36. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
37. Meredith, R. W., Gatesy, J., Murphy, W. J., Ryder, O. A. & Springer, M. S. Molecular decay of the tooth gene *Enamelin* (*ENAM*) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet.* **5**, e1000634 (2009).
38. Kondrashov, F. A. & Koonin, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet.* **20**, 287–290 (2004).
39. Aury, J.-M. *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178 (2006).
40. Gout, J.-F., Kahn, D., Duret, L. & Paramecium Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* **6**, e1000944 (2010).
41. Yanai, I., Peshkin, L., Jorgensen, P. & Kirschner, M. W. Mapping gene expression in two *Xenopus* species: evolutionary constraints and developmental flexibility. *Dev. Cell* **20**, 483–496 (2011).
42. Langley, A. R., Smith, J. C., Stemple, D. L. & Harvey, S. A. New insights into the maternal to zygotic transition. *Development* **141**, 3834–3841 (2014).
43. Marcet-Houben, M. & Gabaldón, T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker’s yeast lineage. *PLoS Biol.* **13**, e1002220 (2015).
44. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).

Supplementary Information is available in the online version of the paper.

Acknowledgements Please see Supplementary Note 15 for funding information.

Author Contributions R.M.H., M.T., D.S.R., G.J.C.V., A.Fuj., A.S., A.M.S., T.Kw., Y.U., A.Fuk., M.K. and H.Og. provided project leadership, with additional project management from Y.M., M.A., Y.Iz., N.U., J.Sh., J.B.W., E.M.M., J.Sc., A.M.Z., P.D.V. and M.I. Y.Iz. and J.R. inbred J strain frogs. A.T., C.H., A.Fuj., J.G., J.C., J.K., J.Sh., T.Mit. and J.B.L. generated genome sequence data. J.A.C., A.M.S., T.Kw., J.J., A.Fuk., M.T. and J.Sc. performed genome assembly and validation. S.T., T.Kw., A.M.S., Y.S., T.T., A.T., A.S. and M.T. generated and analysed the transcriptome data. A.M.S., T.Kw., S.J.v.H. and S.S. generated the annotations. Manual validation of annotation was done by H.Og., S.T., A.Fuk., A.S., M.K., H.Oc., T.T., T.Mic., M.W., T.Ki., Y.O., S.Ma., Y.H., T.N., Y.Y., J.F., K.B., V.L., D.W.D., M.T. and K.K. K.M., A.M.S. and R.H. generated the Hymenochirus transcriptome data. A.M.S. performed the phylogenetic analysis, with contributions from S.V.M. and U.H. M.W., A.Fuk.,

S.Ma., Y.U., Y.M. and M.T. performed the chromosome structure analysis. A.M.S., A.H., O.S., J.C. and Y.U. studied the transposable elements. BAC-FISH was performed by Y.U., A.Fuk., M.K., A.T., S.T., H.Og., H.Oc., Y.K., T.T., T.M., M.W., T.Ki., Y.O., Y.H., T.S.Y., C.T., T.N., A.S., Y.M., N.U., M.A., Y.Iz., A.Fuj. and M.T. I.Q., S.H., N.P. and J.St. generated and analysed the chromatin conformation capture data and their use in long-range scaffolding. H.Og. and H.Oc. performed the transgenic enhancer analysis. S.J.v.H., G.G., S.S.P., I.v.K., O.B., R.L., and G.J.C.V. generated and analysed the epigenetic data. A.S., A.M.S., T.Ki., M.K., M.T., Y.O., T.T., A.Fuk., M.W., T.Mic., D.W.H., T.N. and L.D. conducted the gene and pathway analysis. D.S.R., A.M.S., T.Kw., R.M.H., M.T., A.S., Y.U., G.J.C.V., M.K., U.H., S.J.v.H., A.Fuk., A.H., O.S., H.Og., T.T., I.Q., J.K., Y.O., S.T., M.W., T.Mic., A.T., H.Oc., T.Ki., S.Maw., Y.S., T.N., Y.Iz. and M.F.F. wrote the paper and supplementary notes, with input from all authors.

Author Information NCBI (LYTH00000000). Sequence Read Archive (SRP071264, SRP070985). NCBI Gene Expression Omnibus (GSE73430, GSE73419, GSE76089, GSE76059, GSE76247). DDBJ/GenBank/EMBL

(AP017316 and AP017317). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.H., (harland@berkeley.edu) M. T. (m_taira@bs.s.u-tokyo.ac.jp) or D.S.R (dsrokhsar@gmail.com).

Reviewer Information *Nature* thanks C. Amemiya, S. Burgess and the other anonymous reviewer(s) for their contribution to the peer review of this work.



This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Notation and terminology. ‘Homoeologous’ chromosomes are anciently orthologous chromosomes that diverged by speciation but were reunited in the same nucleus by a polyploidization event. They are a special case of paralogues. Homoeologous genes are sometimes called ‘alloalleles’ to emphasize their role as alternate forms of a gene, but since homoeologues are unlinked and assort independently, we do not use this terminology. Similarly, loss of homoeologous genes is sometimes referred to as ‘diploidization’. We prefer the simpler and more descriptive term ‘gene loss’. Note that an allotetraploid such as *Xenopus laevis* has two related subgenomes, but these subgenomes are each transmitted to progeny via conventional disomic inheritance. So immediately after allotetraploidization, the new species is already genetically diploid. This is clearly the case for *X. laevis*, since we find no evidence for recombination between homoeologous chromosomes, which would create new sequences with mixed ‘L’ and ‘S’ type transposable elements.

Sequencing and assembly. DNA was extracted from the blood of a single female from the inbred J-strain for whole-genome shotgun sequencing. We generated 4.6 billion paired-end Illumina reads from a range of inserts and used Sanger dideoxy sequencing to obtain fosmid- and bacterial artificial chromosome (BAC)-end pairs and full BAC sequences. We used meraculous⁴⁵ as the primary genome assembler. See supplementary notes for more detailed information.

Chromosome scale organization. We identified 798 BACs containing genes of interest distributed across the *Xenopus* genome and performed fluorescence FISH to assign these BACs to specific chromosomes based on Hoechst 33258-stained late-replication banding patterns (Supplementary Table 1). Tethered chromatin conformation capture (TCC)⁴⁶ and in vitro chromatin conformation capture⁴⁷ were performed as previously described, and assembled with HiRise⁴⁷.

Characterization of sex locus. Sex determination in *X. laevis* follows a female heterogametic ZZ/ZW system⁴⁸. We fully sequenced BAC clones representing both W and Z haplotypes, and identified both W- and Z-specific sequences (Extended Data Fig. 2a). The existence of the Z-specific sequence was unexpected and therefore verified by PCR analysis using specific primer sets and DNA from gynogenetic frogs having either W or Z loci.

Gene annotation. We made use of extensive previously generated transcriptome data for *X. laevis* and *X. tropicalis*, including 697,015 *X. laevis* EST sequences⁴⁹. In addition, more than 1 billion RNA-seq reads were generated for this project from 14 oocyte/developmental stages and 14 adult tissues from J-strain *X. laevis* (Supplementary Note 4). These data were combined with homology and *ab initio* predictions using the Joint Genome Institute’s integrated gene call pipeline (see Supplementary Notes 4 and 8 for more details).

Characterization of subgenome-specific transposable elements. We found subgenome-specific repeats using a RepeatMasker⁵⁰ result. The repeats were used to reconstruct full-length subgenome specific transposon sequences. The specific transposons, Xl-TpL_harb, Xl-TpS_harb and Xl-TpS_mar, were classified on the basis of their target site sequence and terminal inverted repeat (TIR) sequences. The coverage lengths of the transposons on each chromosome were calculated from the results of BLASTN search ($E < 10^{-5}$) using the consensus sequences of the transposons as queries. The chromosomal distribution of the Xl-TpS_mar was revealed by a FISH analysis (Supplementary Note 7.4).

Phylogeny, divergence time, and evolutionary rates. We used *Hymenochirus boettgeri*, *Pipa carvalhoi* and *Rana pipiens* sequences as outgroups to estimate the evolutionary rate of duplicated genes in *X. laevis* and their relationship to *X. tropicalis*. See Supplementary Notes 7 and 8 for more detail.

Deletions and pseudogenes. Pseudogene sequences contain various defects including premature stop codons, frameshifts, disrupted splicing, and/or partial coding deletions. 985 pseudogenes were identified among 1,531 ‘2-1-2 regions’, with the others deleted or rendered unidentifiable by mutation. 368 out of 985 could be timed, based on the accumulation of non-synonymous and synonymous substitution between a pseudogene, its homoeologue and its orthologue in *X. tropicalis*, providing a time since the loss of constraint for each pseudogene³⁷. See Supplementary Note 9 for additional details.

Functional annotation of genes. We used several bioinformatic methods and high-throughput datasets to assign functional annotations to *Xenopus* genes.

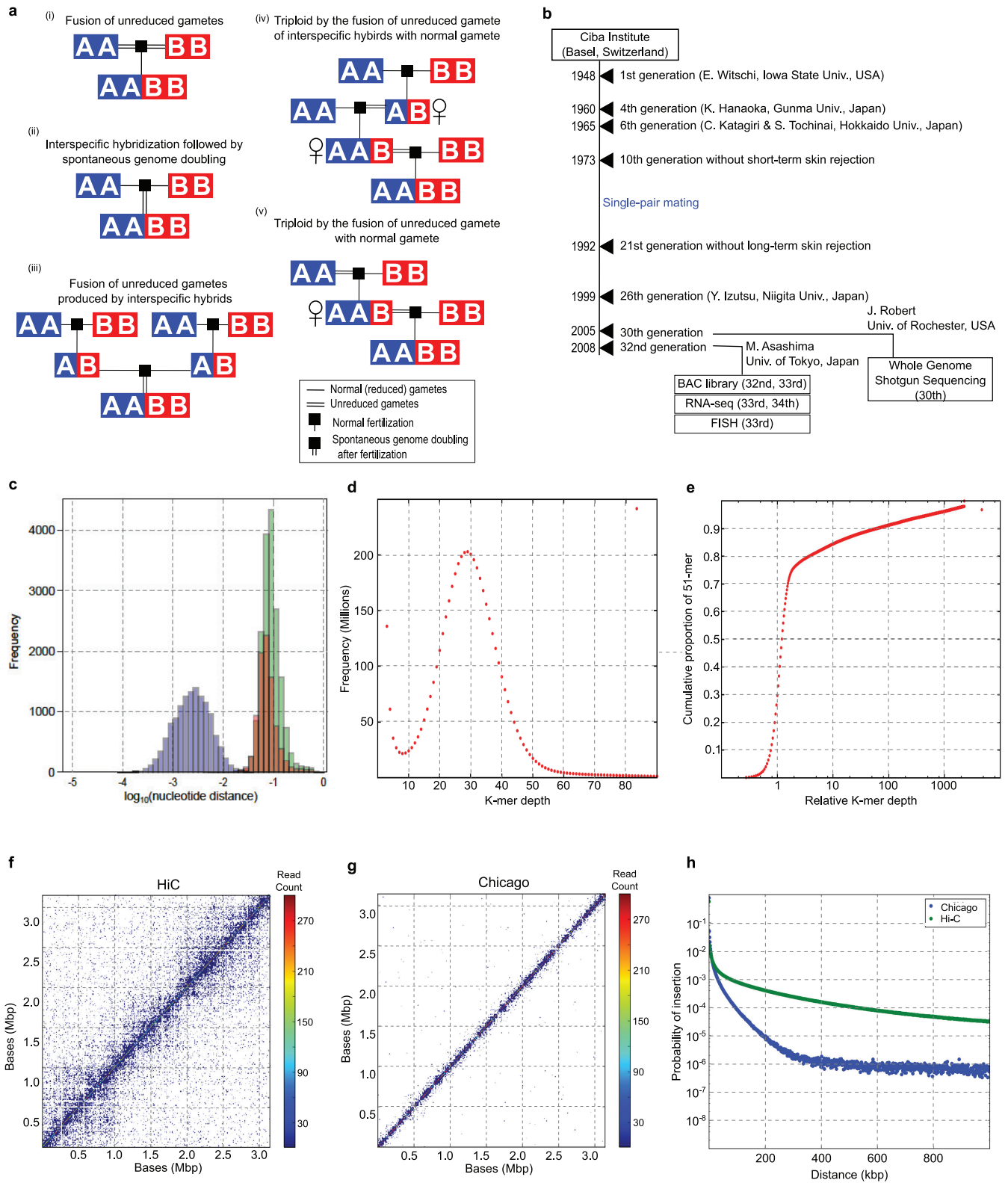
Protein domains were assigned using InterPro (including PFAM and Panther)⁵¹ and KEGG⁵². Gene Ontology was assigned using InterPro2Go⁵¹. We identified genes that encode mitochondrial proteins by mapping the MitoCarta⁵³ database from mouse to the most recent *X. tropicalis* proteome. *Xenopus* genes associated with germ plasm were manually curated using the extensive *Xenopus* literature (Supplementary Note 13).

Gene expression. We analysed transcriptome data generated for 14 oocyte/developmental stages and 14 adult tissues in duplicate except for oocyte stages (see Supplementary Note 4). Expression levels were measured by mapping paired-end RNA-seq reads to predicted full length cDNA and reporting transcripts per one million mapped reads (TPM). We consider the limit of detectable expression to be TPM >0.5. Co-expression modules were defined by weighted gene correlation network analysis (WGCNA) clustering⁵⁴ (Supplementary Note 12).

Epigenetic analysis. We determined DNA methylation levels (DNAm) by whole genome bisulfite sequencing and used ChIP-seq to generate profiles of the promoter mark histone H3 lysine 4 trimethylation (H3K4me3), the transcription elongation mark H3K36me3, as well as RNA polymerase II (RNAPII) and the enhancer-associated co-activator p300. To test which regulatory features would contribute most to the L versus S expression differences, we applied a Random Forest machine learning algorithm to analyse differential expression between the L and S homoeologues (See Supplementary Note 14).

Data availability. The XENLA v9.1 genome assembly and annotation are deposited at NCBI (accession LYTH000000000). The DNA read libraries of *X. laevis* and *X. borealis* were deposited at the Sequence Read Archive under accessions SRP071264 and SRP070985, respectively. Datasets of the *X. laevis* RNA-seq short reads were deposited in NCBI Gene Expression Omnibus (accession number GSE73430 for stages, GSE73419 for tissues). Datasets of the *Hymenochirus* RNA-seq short reads were deposited in NCBI GEO (accession number GSE76089). The epigenetic data have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO Series accession numbers GSE76059 for ChIP-seq. MethylC-seq data are accessible through GEO Series accession number GSE76247. The sequence data from BAC and fosmid clones have been deposited to DDBJ/GenBank/EMBL under the accession numbers: (i) GA131508–GA227532, GA228275–GA244139, GA244852–GA274229, GA274976–GA275712, GA277157–GA344957, GA345673–GA350926 and GA351685–GA393223 for the XLB1 end-sequences; (ii) GA720358–GA756840 for the XLB2 end-sequences; (iii) GA756841–GA867435 for the XLFIC end-sequences and (iv) AP012997–AP013026, AP014660–AP014679, AP017316 and AP017317 for the finished BAC/fosmid sequences.

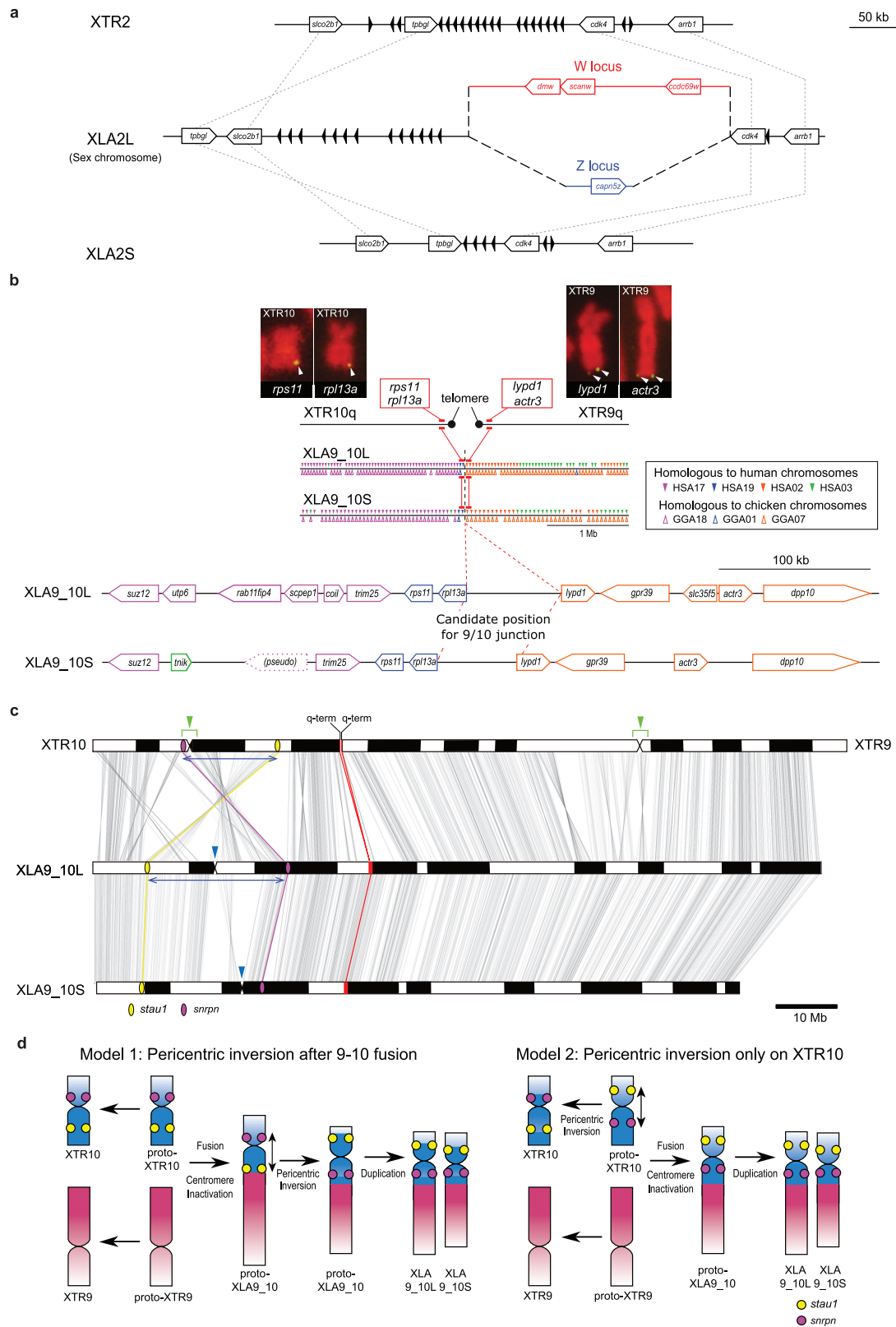
45. Chapman, J. A. *et al.* Meraculous: de novo genome assembly with short paired-end reads. *PLoS One* **6**, e23501 (2011).
46. Chen, L. *et al.* Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2011).
47. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
48. Chang, C. Y. & Witschi, E. Genic control and hormonal reversal of sex differentiation in *Xenopus*. *Proc. Soc. Exp. Biol. Med.* **93**, 140–144 (1956).
49. Gilchrist, M. J. From expression cloning to gene modeling: the development of *Xenopus* gene sequence resources. *Genesis* **50**, 143–154 (2012).
50. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
51. Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221 (2015).
52. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205 (2014).
53. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).
54. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
55. Edwards, N. S. & Murray, A. W. Identification of *Xenopus* CENP-A and an associated centromeric DNA repeat. *Mol. Biol. Cell* **16**, 1800–1810 (2005).
56. McLysaght, A. *et al.* Ohnologs are overrepresented in pathogenic copy number mutations. *Proc. Natl. Acad. Sci. USA* **111**, 361–366 (2014).
57. Tan, M. H. *et al.* RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome Res.* **23**, 201–216 (2013).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Allotetraploidy and assembly. a–e, Scenarios for allotetraploid formation from distinct ancestral diploid species A and B. Horizontal single lines indicate normal gametes, horizontal double lines indicate unreduced gametes; black square represents fertilization; vertical double lines indicate spontaneous (somatic) genome doubling. **a,** (i) Fusion of unreduced gametes from species A and B. (ii) Interspecific hybridization followed by spontaneous doubling. (iii) Fusion of unreduced gametes produced by interspecific hybrids. (iv) Interspecific hybrids produce unreduced gametes, which fuse with normal gametes from species A. The resulting triploid again produces unreduced gametes, which fuse with normal gametes from species B. (v) Unreduced gamete from species A fuses with normal gamete from species B. The resulting AAB triploid produces unreduced gametes that are fertilized by normal gametes species B. See Supplementary Note 1.1 for a more detailed discussion. **b,** History of the J strain. See Supplementary Note 2.1 for details. The years of events and generation numbers (such as frog transfer to another institute, establishment of homozygosity, construction of materials) are indicated in the scheme. Generation numbers are estimates due to loss of old breeding records. **c,** The nucleotide distance of orthologues (green), homoeologues (red) and alleles (blue) is discussed in Supplementary

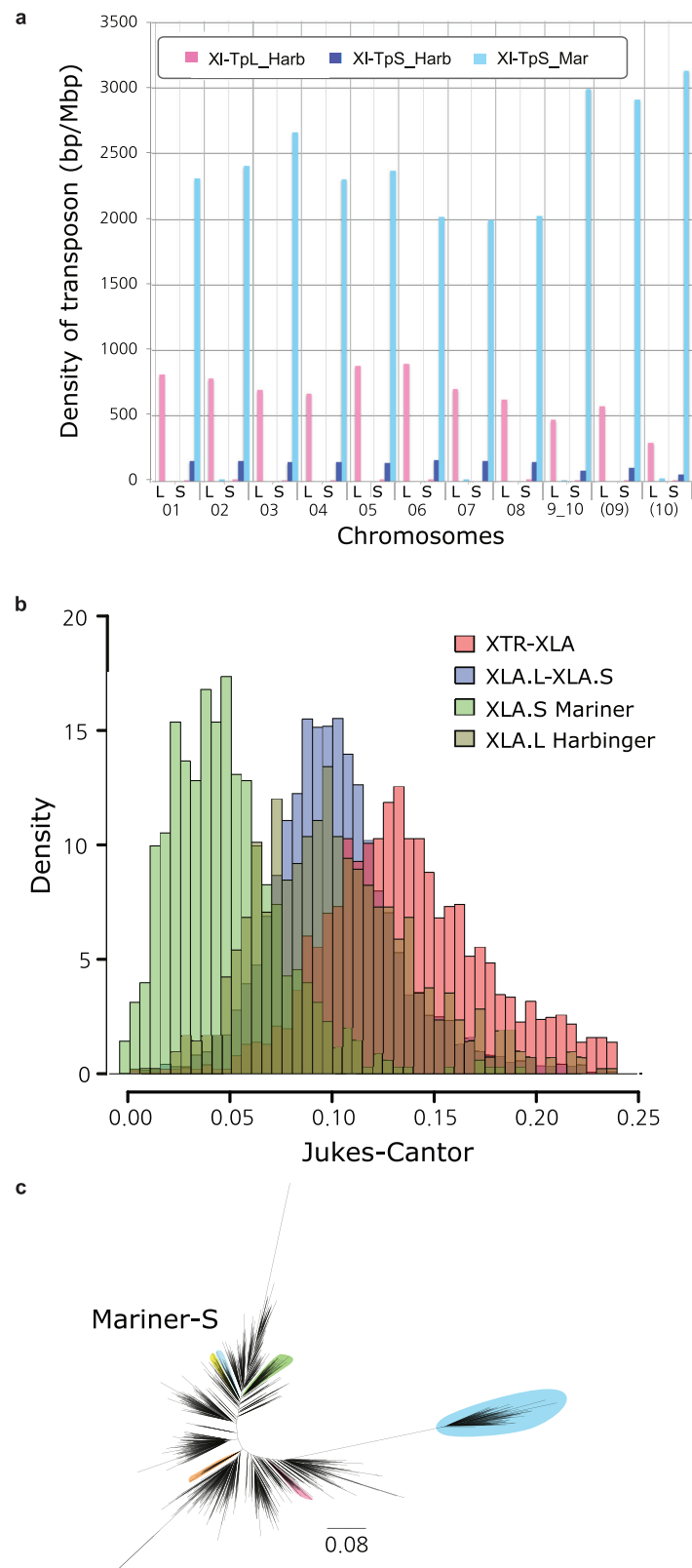
Note 8.7. The distances are shown on a log scale to differentiate between the distributions. **d,** Frequency histogram showing the number of 51-mers with specified count in the shotgun dataset. The prominent peak implies that each genomic locus is sampled $29\times$ in 51-mers. Note the absence of a feature at twice this depth, indicating that homoeologous features with high identity are rare. **e,** Cumulative proportion of 51-mers as a function of relative depth (that is, depth/29). Relative depth provides an estimate of genomic copy number. The rapid rise at relative depth 1 implies that 70–75% of the *X. laevis* genome is a single copy with respect to 51-mers. The remainder of the genome is primarily concentrated in repetitive sequences with copy number > 100 . Note logarithmic scale. **f,** The contact map of 85,260 TCC read pairs for JGIv72.000090484.chr4S. Read pairs were binned at 10-kb intervals. For each read pair, the forward and reverse reads map with a map quality score of at least 20. **g,** The contact map of 85,260 Chicago read pairs for JGIv72.000090484.chr4S, a 3.1-Mb scaffold in the XENLA_JGI_v72 assembly. **h,** The insert distribution of TCC and Chicago read pairs that map to the same scaffold of XENLA_JGI_v72 with a map quality score of at least 20. The *x* axis is the read pair separation distance. The *y* axis is the counts for that bin divided by the total number of reads. The bins are 1 kb.



Extended Data Figure 2 | See next page for caption.

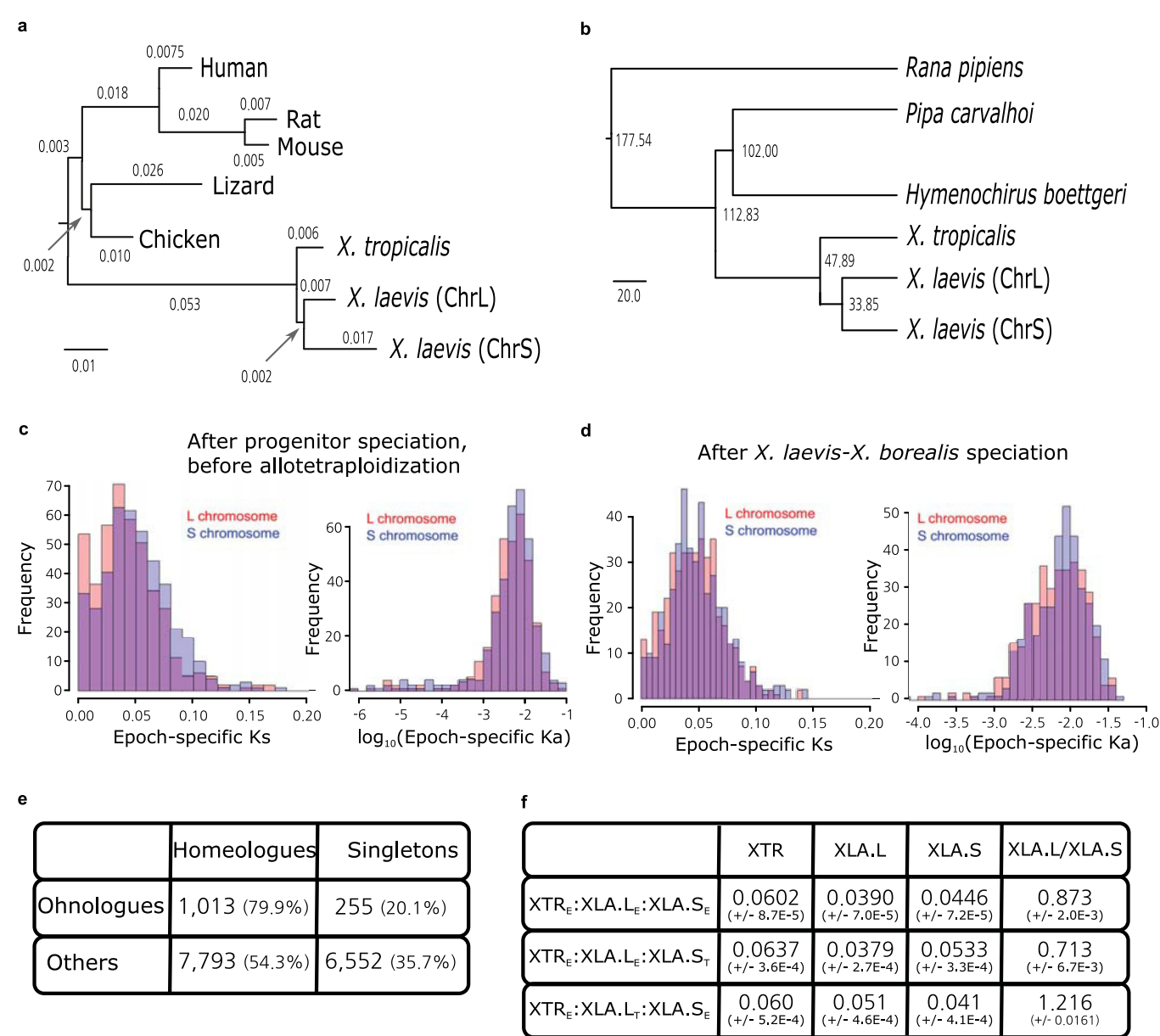
Extended Data Figure 2 | Chromosome structure. **a**, Structure of the sex chromosome of *X. laevis* (XLA2L) and comparison with XLA2S and XTR2. The W version of XLA2L harbours a W-specific sequence containing the female sex-determining gene *dmw* (red) while Z has a different Z-specific sequence (blue). Pentagon arrows and black triangles indicate genes and olfactory receptor genes, respectively. Their tips correspond to their 3'-ends. **b**, Alignment of the q-terminal regions of XTR9 and 10 with corresponding regions of XLA9_10L and XLA9_10S. Genes near the q-terminal regions of XTR 9 and XTR10 were missing in the *X. tropicalis* genome assembly v9, but *rps11*, *rpl13a*, *lypd1* and *actr3* were expected to be located there based on the synteny with human chromosomes, and then verified by cDNA FISH (upper panels). Small triangles on XLA9_10L and S indicate the distribution of gene models showing both identity and coverage greater than 30%, against the human and chicken peptide sequences from Ensembl, in the region ± 2 Mb from the prospective 9/10 junction. HSA, human chromosome; GGA, chicken chromosome. The magnified view represents syntenic genes to scale with colours corresponding to human genes. **c**, The orders of orthologous genes across XTR9, XTR10, XLA9_10L and XLA9_10S. Green arrowheads: positions of centromeres in XTR9 and 10 predicted by examination of the cytogenetic chromosome length ratio of p versus q arms¹⁵. Blue

arrowheads: positions of centromere repeats, frog centromeric repeat-1 (ref. 55), in XLA9_10L and S. Magenta and yellow ellipses, chromosomal locations of *snrpn* (magenta) and *stau1* (yellow) from *X. tropicalis* v9 and *X. laevis* v9.1 assemblies. Red ellipses, chromosomal locations of four genes, *rps11*, *rpl13a*, *lypd1* and *actr3*. XTR9 is inverted to facilitate comparison. Blue bidirectional arrows indicate the homologous regions where pericentric inversions may have occurred on proto-chromosomes (see Extended Data Fig. 2d). **d**, Schematic representation for the two hypothetical processes of chromosomal rearrangements (fusion and inversion) that occurred between the hypothetical proto-XTR9 and 10 to produce proto-XLA9_10, and eventually XLA9_10L and S. The process of chromosome rearrangements is explained parsimoniously in two different ways (left and right panels), starting from proto-XTR9 and 10. Actual and hypothetical ancestral chromosomal locations of *snrpn* and *stau1* are shown by magenta and yellow circles, respectively. Note that the chromosomal locations of these genes on the proto-XTR10 differ between the two models. Chromosome segments homologous to XTR9 and XTR10 are shown in red and blue, respectively. XTR9 is inverted to facilitate comparison. Bidirectional arrows indicate the regions where pericentric inversions may have occurred. Black arrows indicate the direction of chromosomal evolution.



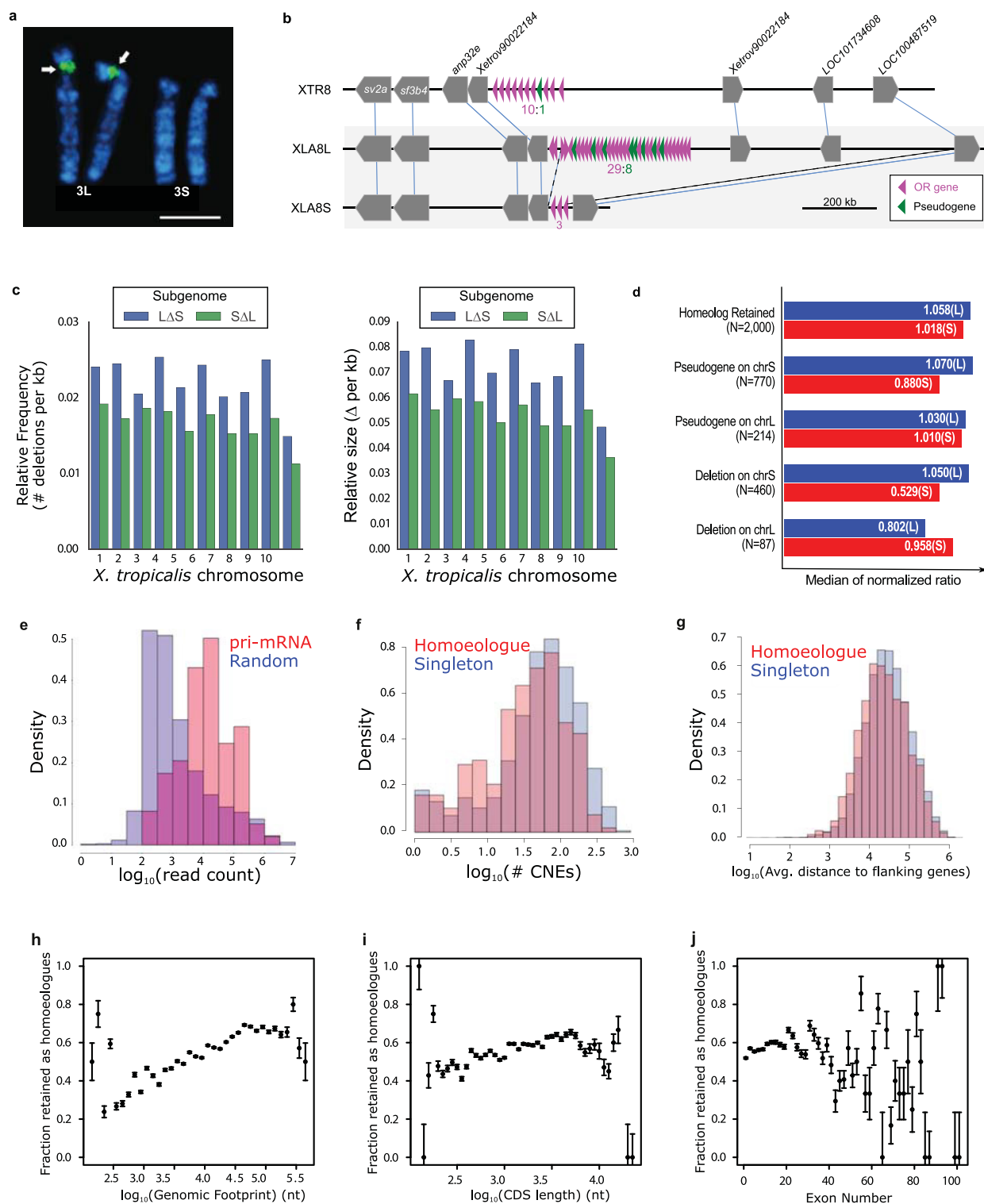
Extended Data Figure 3 | Transposons. **a**, Density of the subgenome-specific transposons on each chromosome (coverage length of transposable element (bp)/chromosome length (Mbp)). The coverage lengths of transposons were calculated from the results of BLASTN search (E-value cutoff 10^{-5}) using the consensus sequences as queries. **b**, Jukes-Cantor distances across non-CpG sites, corrected as in Supplementary Note 7.5. Distances between *X. tropicalis* and *X. laevis* transposons consensus sequences are shown. The *X. laevis*-specific transposon

differences are each individual transposon sequence against the consensus sequence for that subfamily. **c**, Phylogenetic tree of XI-TpS_mar transposon expansions in the *X. laevis* genome, built using Jukes-Cantor corrected distances (Supplementary Note 7.5). Sub-clusters with enough members to determine accurate timings are highlighted. The scale bar represents the corrected Jukes-Cantor distance of 0.08 substitutions per site.



Extended Data Figure 4 | Phylogeny. **a**, Phylogenetic tree of pan-vertebrate conserved non-coding elements (pvcNEs), rooted by elephant shark. Alignments were done by MUSCLE, and the maximum-likelihood tree was built by PhyML. Branch length scale shown at the bottom. The difference in branch lengths of tetrapods follows the same topology as the protein-coding tree (Fig. 2b). **b**, Complete phylogenetic tree from Fig. 2a, with divergence times computed by r8s. **c**, Distribution of synonymous and non-synonymous rates K_s and K_a on specific subgenomes during the time between L and S speciation, before *X. laevis* and *X. borealis* speciation. We find accelerated mutations rates between T2 and T3 in K_s and K_a ($P = 1.4 \times 10^{-5}$ (left), 8.6×10^{-3} (right)). **d**, Distribution of K_s and K_a on specific subgenomes during the time after *X. laevis* and *X. borealis* speciation. We do not find significantly accelerated substitution rates ($P = 0.10$ (left) and $P = 0.03$ (right)). **e**, Table showing the number of homeologues and singletons identified as homeologues from the ancient vertebrate duplication (or ohnologues as they were historically

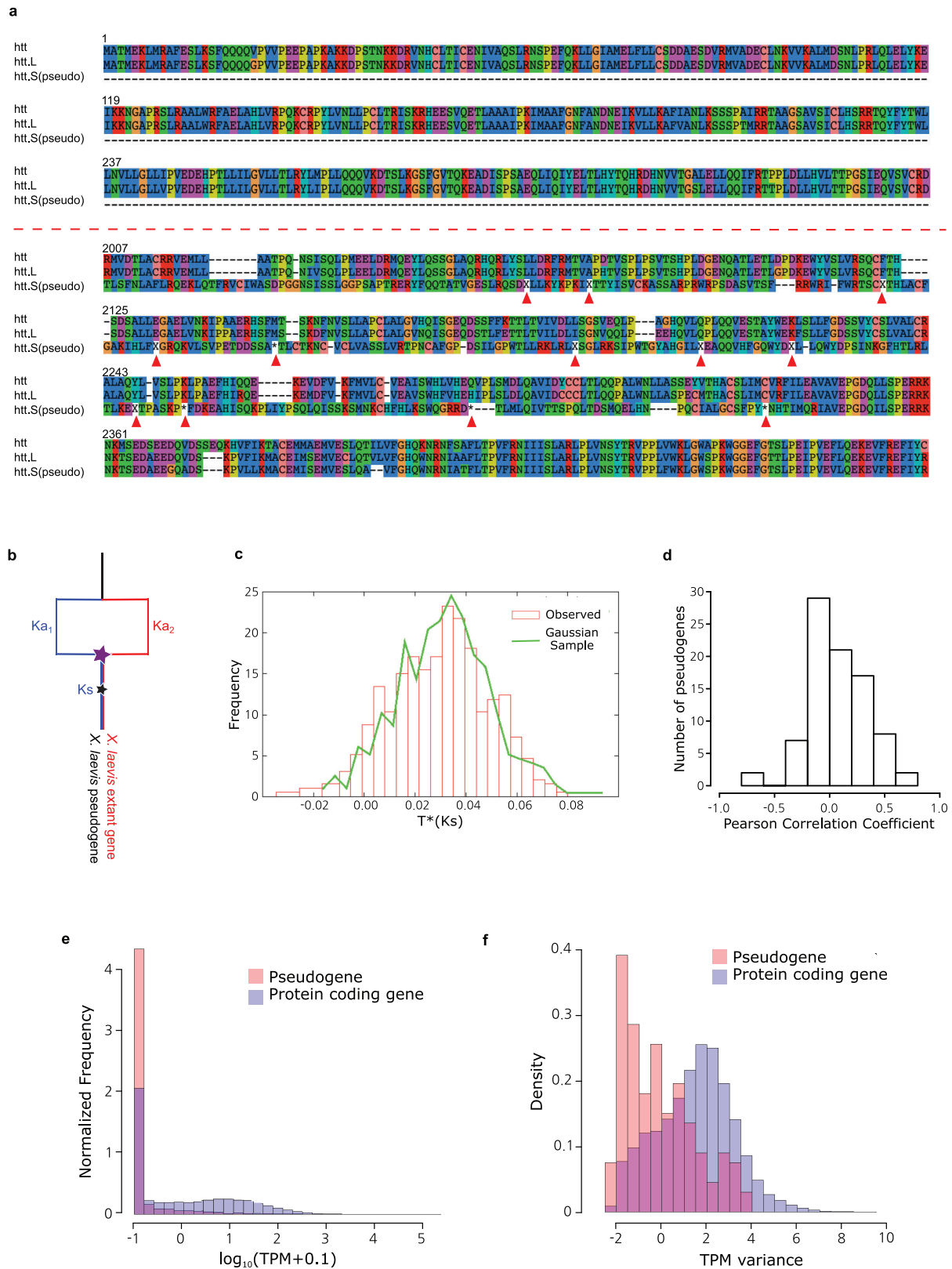
called)⁵⁶, 79.9% of ohnologues retain both copies in *X. laevis* today, significantly more than the 54.3% of the rest of the genome after excluding ohnologues (χ^2 test $P = 4.44 \times 10^{-69}$). **f**, Table showing the branch lengths of bootstrapped maximum likelihood trees described in Supplementary Note 12.5. The columns refer to the *X. tropicalis* (XTR), L chromosome of *X. laevis* (XLA.L), S chromosome of *X. laevis* (XLA.S) and XLA.L/XLA.S branch lengths respectively. The first row shows triplets where all genes show expression, the second row shows triplets where L is a thanogene, and the third row shows triplets where S is a thanogene. The L branch length is significantly smaller when all genes are expressed, or when S is a thanogene (Wilcoxon signed-rank test, $P = 1.7 \times 10^{-216}$ and 6.4×10^{-212} respectively). The S branch length is smaller when L is a thanogene ($P = 2.4 \times 10^{-223}$). The ratio of branch lengths (L/S) is significantly different for either L or S thanogene datasets compared to when all genes are expressed ($P = 3.55 \times 10^{-214}$ and 7.48×10^{-220} respectively). The ratio is also different between the two thanogene datasets ($P = 1.79 \times 10^{-217}$).



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Structural evolution. **a**, Chromosomal locations of the 45S pre-ribosomal RNA gene (*rna45s*), which encodes a precursor RNA for 18S, 5.8S and 28S rRNAs, was determined using pHr21Ab (5.8-kb for the 5' portion) and pHr14E3 (7.3-kb for the 3' portion) fragments as FISH probes. DNA fragments used for the probes were provided by National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, and labelled with biotin-16-dUTP (Roche Diagnostics) by nick translation. After hybridization, the slides were incubated with FITC-avidin (Vector Laboratories). Hybridization signals (arrows) were detected to the short arm of XLA3L, but not XLA3S. Scale bar, 5 μ m. **b**, A large deletion including an olfactory receptor gene (*or*) cluster. Schematic structures of *or* gene clusters and adjacent genes on the 8th chromosomes of *X. tropicalis* (XTR8) and *X. laevis* (XLA8L and XLA8S). Chromosomal locations: XTR8: 107,524,547–108,927,581; XLA8L: 105,062,063–106,610,199; XLA8S: 91,630,596–92,060,451. Horizontal bars, genomic DNA sequences; triangles, genes. Outside of *or* gene cluster, only representative genes are shown. The size of the triangle is to scale. The orientation of triangles indicates 5' to 3' direction of genes. Thin lines connect orthologous/homoeologous genes. Magenta triangles, *or* genes; green triangles, pseudogenes (point-mutated or truncated *or* genes). The number of *or* genes is shown underneath gene clusters. Dotted lines, a deleted region in XLA8S compared to XLA8L. The centromere is located on the left side and the telomere is on the right. **c**, The relative frequency (left panel) and size (right panel) of genomic regions deleted in the S (blue) and L (green) chromosomes respectively. Both subgenomes experienced sequence loss through deletions, but the deletions on the S subgenome are larger and have been more frequent. Deletions were called based on the progressive Cactus sequence alignment between the *X. laevis* L and S subgenomes and the *X. tropicalis* genome. Chromosome 9_10 of *X. laevis* was split into 9 and 10 on the basis of alignment with the *X. tropicalis* chromosomes. Sequences from L that were not present on S, but could at least partially be identified in *X. tropicalis*, and consisted of gaps for no more than 25% of their length, were called as deleted regions in S. The same procedure was followed for deleted regions in L. **d**, Identification of triplet loci is described in Supplementary Note 8.1. Loci were classified into groups based on the presence of gene 2 in both *X. laevis* subgenomes (homoeologue retained), versus those that had a pseudogene in the middle (pseudogene) or no remnant of the middle gene as assessed by Exonerate (deletion). To normalize the intergenic lengths, we divided the

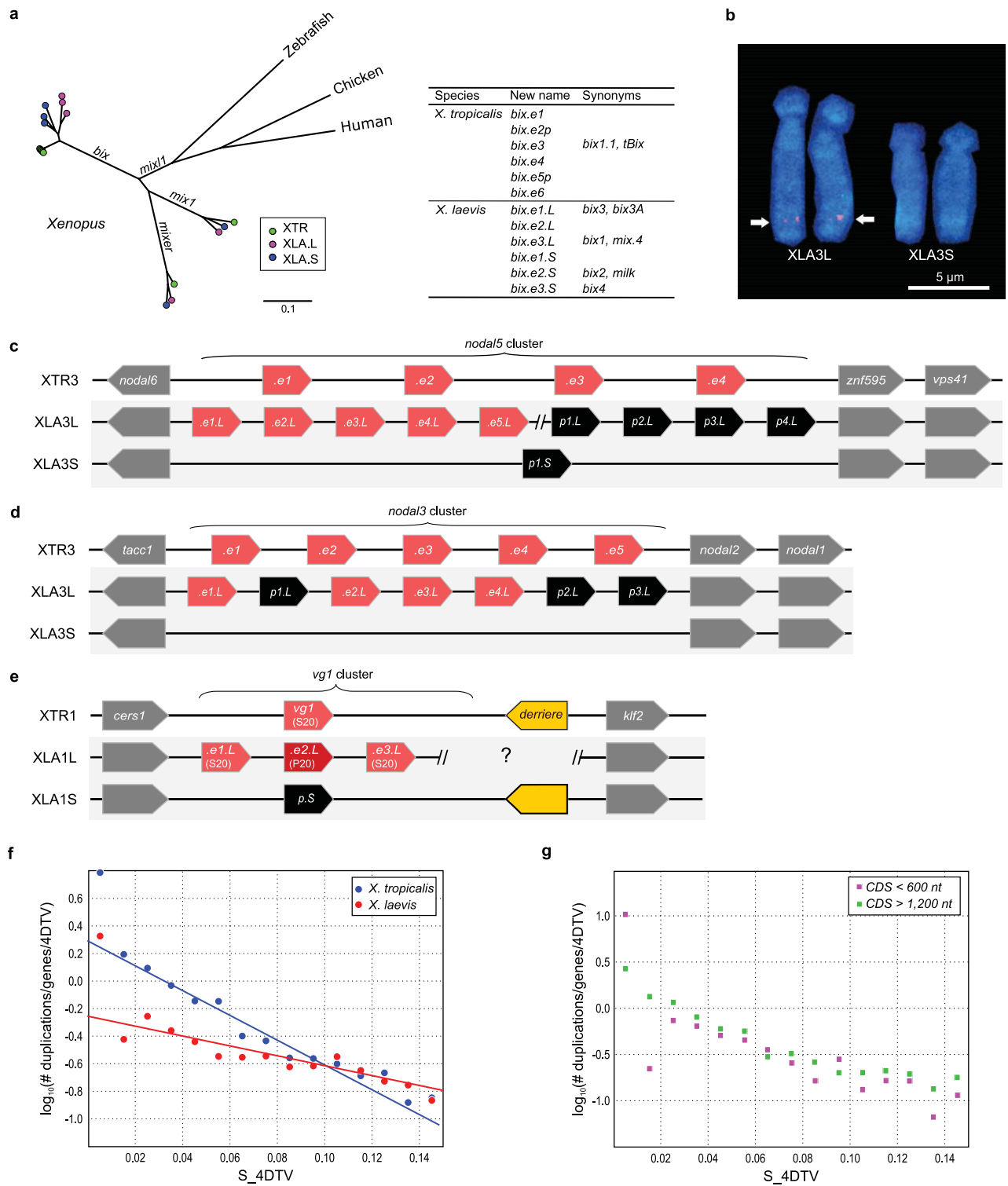
nucleotide distance between genes 1 and 3 in either *X. laevis* subgenome by the orthologous distance in *X. tropicalis*. The median of the normalized ratio distribution is plotted on the bar chart. On average, S deletions appear to be larger than L deletions (52.9% versus 80.2% of the size of the orthologous *X. tropicalis* region, respectively). **e**, The number of RNA-seq reads aligning ± 1 kb of precursor miRNA loci (red) was compared to the read count for 10,000 random unannotated 2.1 kb regions of the genome (blue). All 83 homoeologous, intergenic miRNA pairs showed alignment within their regions, as opposed to 4,127 out of 10,000 (41.27%) of the randomly chosen intergenic sequences. The putative primary-miRNA loci also have a higher read count than the expressed randomly chosen regions (Wilcoxon signed-rank test, $P = 1.4 \times 10^{-38}$). **f**, The Cactus alignment was parsed to identify flanking CNE around each *X. tropicalis* gene. The number of CNEs > 50 bp in length for singletons is shown in red, homoeologues in blue. Kolmogorov-Smirnov test $P = 10^{-11}$. **g**, The average distance to the nearest gene was computed for each chromosomal locus in *X. tropicalis*. The average intergenic distance for those with a single *X. laevis* gene is shown in red, those with two shown in blue. Wilcoxon signed-rank test ($P = 9.8 \times 10^{-24}$). **h**, The distribution of gene retention by genomic footprint of the *X. tropicalis* orthologue. We define genomic footprint as the genomic distance from the start signal of the coding sequence (CDS) to the stop signal, including introns. The *x* axis shows \log_{10} (genomic footprint), the *y* axis the retention rate of each bin. The error bars are the standard deviation of the total divided by the number of genes in each bin. We tested for significant differences in length between homoeologues and singletons by a Wilcoxon signed-rank test ($P = 2.4 \times 10^{-96}$). **i**, The distribution of gene retention by CDS length of the *X. tropicalis* orthologue. The *x* axis shows \log_{10} (CDS length), the *y* axis the retention rate of each bin. The error bars are the standard deviation of the total divided by the number of genes in each bin. We tested for significant differences in length between homoeologues and singletons by a Wilcoxon signed-rank test ($P = 1.7 \times 10^{-21}$). **j**, The distribution of gene retention by exon number of the *X. tropicalis* orthologue. The *x* axis shows number of exons; the *y* axis the retention rate of each bin. The error bars are the standard deviation of the total divided by the number of genes in each bin. We tested for significant differences in length between homoeologues and singletons by a Wilcoxon signed-rank test ($P = 3.2 \times 10^{-8}$).



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Pseudogenes. **a**, Illustration of *htt.S* pseudogene alignment to *X. tropicalis htt* and the extant *X. laevis htt.L*, translated to amino acids. The amino acid position is shown at the beginning of each line. Missing codons are marked by dashes. Frameshifts and premature stops are marked by X and *, respectively (and pointed to with red arrows). The first exon of the pseudogene is completely missing from the S chromosome (top). The characteristic poly-Q region is maintained by both *htt* and *htt.L*. An exon with conservation in the pseudogene (bottom), illustrating that despite many frameshifts, premature stops, the lack of a proper start and insertions of new sequence, we identify many codons in the pseudogene that occur in large conserved blocks. **b**, Illustration of our model to compute pseudogene ages. The star represents the point of nonfunctionalization for a locus that is currently a pseudogene. We assume the expected rate of nonsynonymous changes can be estimated by the K_a of the extant gene and *X. tropicalis*. We then compare the K_s and K_a of the

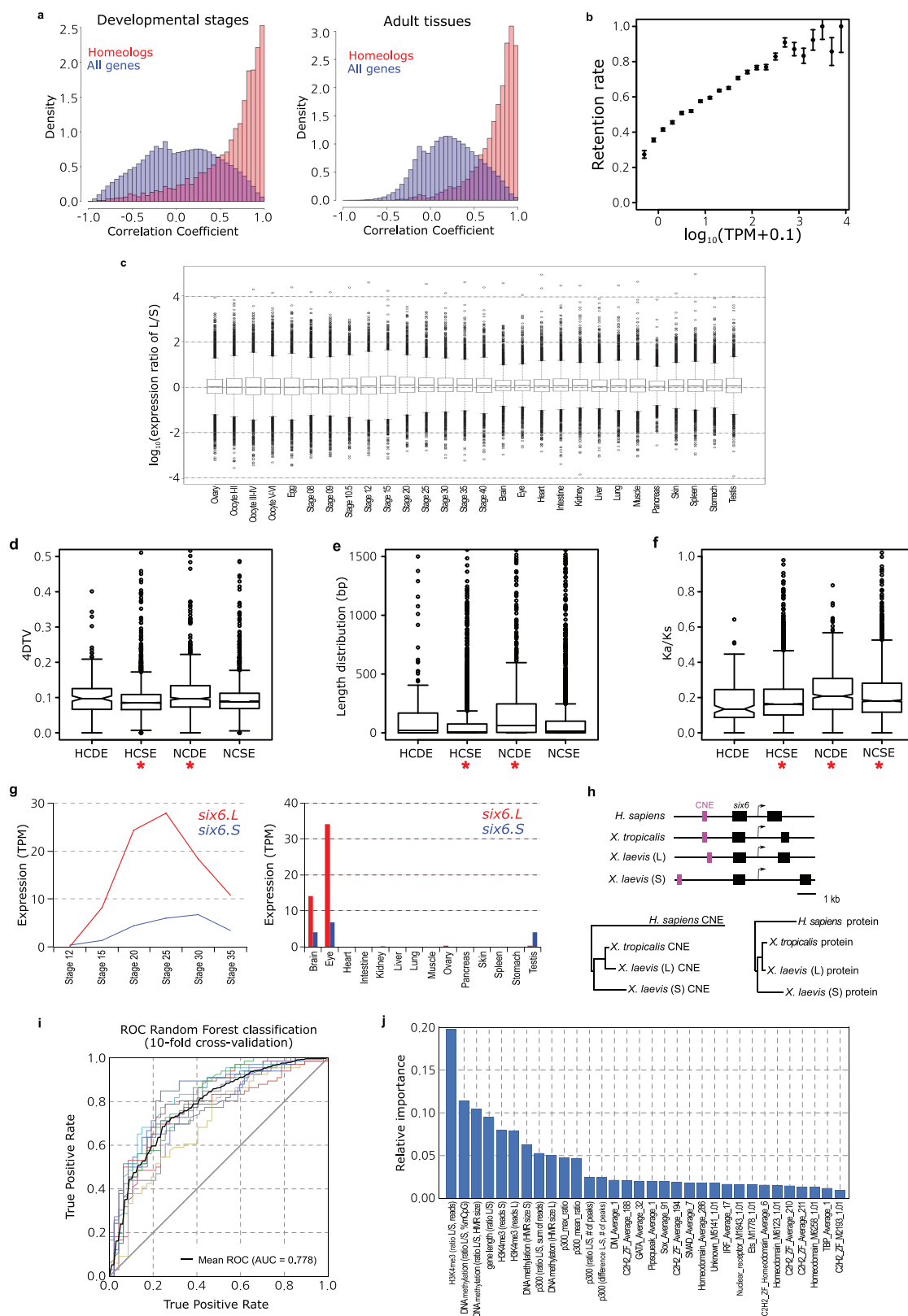
pseudogene sequence to estimate the time of nonfunctionalization. See Supplementary Note 9 for a more detailed discussion. **c**, Estimated epochs of pseudogenization for 430 genes are indistinguishable from a burst of pseudogenization >10 Ma ($K_s > 0.03$). See Supplementary Note 9 for a more detailed discussion. **d**, Correlation of pseudogene expression with its extant homoeologue. The little expression seen in pseudogenes tends to be uncorrelated with the extant homoeologue. **e**, Histogram of pseudogene expression values across all 28 tissues and developmental stages (red) compared to all extant genes (blue). The pseudogenes are rarely expressed and tend to be expressed at lower levels than extant protein-coding genes. **f**, Histograms of expression variance of pseudogenes (red) compared to extant genes (blue). The small amount of pseudogene expression observed does not tend to vary across tissues and developmental stages in the same way that extant genes do.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Tandem duplications. **a**, Phylogenetic trees of the *mix/bix* cluster. Nucleotide sequences were aligned using MUSCLE and a phylogenetic diagram was generated by the ML method with 1,000 bootstraps (MEGA6). Circles with different colours represent *X. laevis* L genes (magenta), *X. laevis* S genes (blue) and *X. tropicalis* genes (green). The table shows the correspondence of *bix* gene names proposed in this study and previously used (synonyms). **b**, FISH analysis showing XLA3S-specific deletion of the *nodal5* gene cluster. One unit of the *nodal5* gene region, including exons, introns and an intergenic region was used as a probe for FISH (counterstained with Hoechst). Arrows indicate the hybridization signals of *nodal5s*. Scale bar, 5 μ m. **c**, Comparison of the *nodal5* gene cluster. Genome sequencing revealed that *nodal5.e1.L~.e5.L* (pink) and *nodal6.L* are clustered. Amplification of *nodal5* gene in XLA3L and loss of this cluster in XLA3S were confirmed. Pseudogenes (*nodal5p1.L~p4.L* and *nodal5p1.S*) are indicated in black. The *nodal5* cluster of *X. tropicalis* does not contain any pseudogene. **d**, The *X. laevis* L chromosome has four complete copies of *nodal3* (*nodal3.e1.L~.e4.L*), whereas the gene cluster is lost from the *X. laevis* S chromosome.

A truncated *nodal3* gene (*nodal3p1.L*) is likely to be a pseudogene and highly degenerate pseudogenes (*nodal3p2.L* and *nodal3p3.L*) also exist on the L chromosome. **e**, Like *nodal3*, *vg1* is lost from the S chromosome although there is a pseudogene (*vg1p.S*). *vg1* is specifically amplified on the *X. laevis* L chromosome (*vg1.e1.L~.e3.L*) in comparison with *X. tropicalis*. An amino acid change (Ser20 to Pro20) in Vg1 protein has been shown to result in functional differences (Supplementary Note 13.9). *vg1* and *derrière* are orthologous to mammalian *gdf1*. **f**, Fraction of all genes duplicated and retained to present epoch per 1 expected 4DTV (fourfold degenerate transversion) at different epochs (semi-log scale). Shown also are linear fits, which would be consistent with constant birth- and death-rate models (first epoch is omitted from both fitted datasets, as is second epoch from *X. laevis*). See Supplementary Note 11 for a more detailed discussion. **g**, Same as **f**, but for 'short genes' (CDS < 600 bp) and 'long genes' (CDS > 1,200 bp) separately. The loss rate of new duplicates appears to be similar. If the extra copy of a newly duplicated gene was lost when the first 100% disabling mutation occurred, we would expect, on average, the longer genes to be lost.

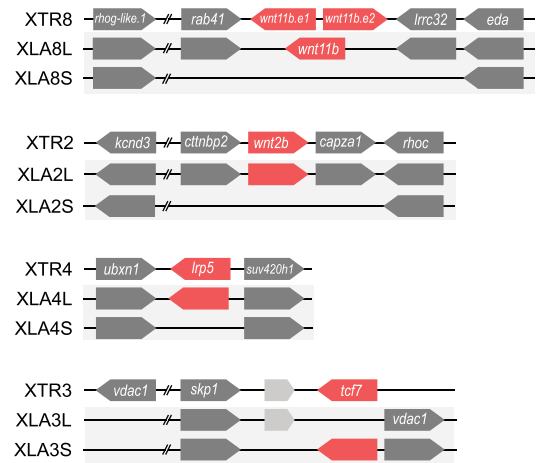
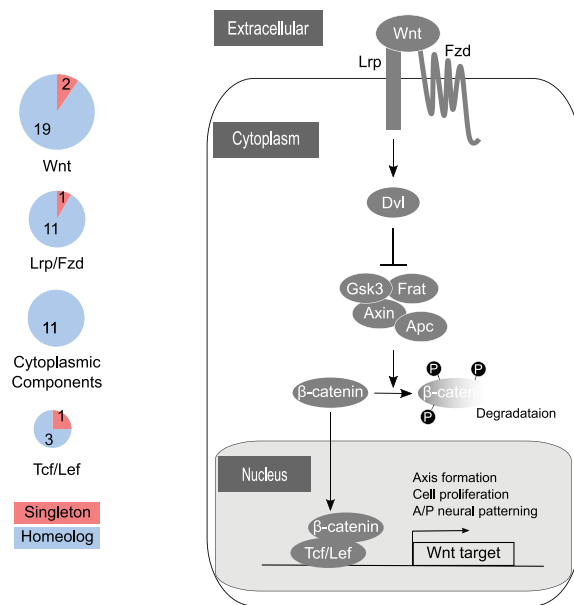


Extended Data Figure 8 | See next page for caption.

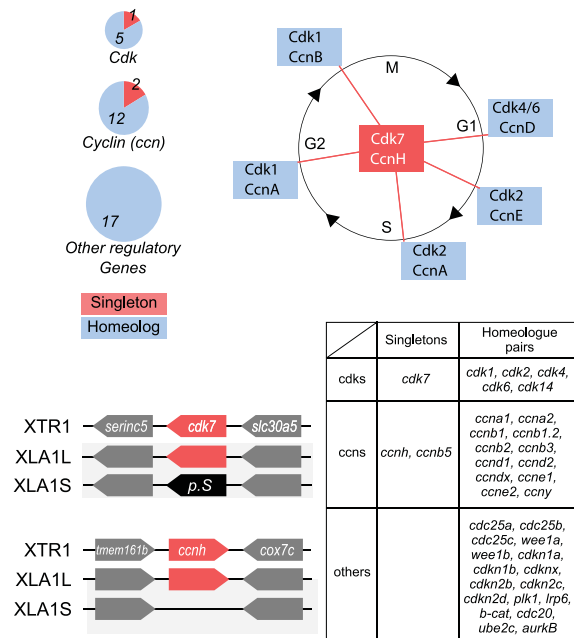
Extended Data Figure 8 | Gene expression analysis. **a**, Pairwise Pearson correlation distributions between homoeologous genes (red) and all genes (blue). Left histogram, stage data; right, adult data. The *x* axis shows the correlation; the *y* axis the percentage of data. The homoeologous genes have a correlation distribution closer to one owing to the fact that these were recently the same locus. *X. laevis* TPM values of 0.5 were lowered to 0. Any gene with no TPM > 0 was removed from analysis. We then added 0.1 to all TPM values and log transformed (\log_{10}) them. **b**, Scatter plot comparing binned genes by their median *X. tropicalis* expression⁵⁷ to the retention rate of their *X. laevis* (co)-orthologues. Error bars are the standard deviation for the whole dataset divided by the square root of the number of genes analysed in a bin. We assessed significance by a Wilcoxon signed-rank test of the homoeologous and singleton distributions, $P = 6.31 \times 10^{-113}$. **c**, Full version of the box plot shown in Fig. 4c. The difference between subgenomes is difficult to see at this magnification, illustrating that many loci deviate from the whole genome median of preferring the L homoeologue. There were some L outliers expressed 10^4 as much as their S homoeologues, whereas no S genes showed such a strong trend. These differences are discussed in more detail in Supplementary Note 12. **d**, Box plot of 4DTV by homoeologue class defined in Supplementary Note 12.4. Significant differences are marked by a red asterisk (Wilcoxon signed-rank test, $P < 10^{-5}$). The high correlation, similar expression (HCSE) group showed lower sequence change than other groups ($P = 3.7 \times 10^{-12}$) and the no correlation, different expression (NCDE) group showed high rates of sequence change ($P = 5.6 \times 10^{-14}$). **e**, Box plot of CDS length difference between *X. laevis* homoeologues by homoeologue class defined in Supplementary Note 12.4. Significant differences are marked by a red asterisk (Wilcoxon signed-rank test, $P < 10^{-5}$). The HCSE group showed smaller CDS length differences than other groups ($P = 2.4 \times 10^{-13}$) and the NCDE group showed large differences in homoeologue CDS length ($P = 2.1 \times 10^{-32}$). **f**, Box plot of K_a/K_s between *X. laevis* homoeologues

by homoeologue class defined in Supplementary Note 12.4. Significant differences are marked by a red asterisk (*t*-test $P < 10^{-5}$). The HCSE group showed lower non-synonymous sequence change than other groups ($P = 8.2 \times 10^{-19}$) and the NCDE and no correlation, similar expression (NCSE) groups showed higher rates of non-synonymous sequence changes ($P = 2.0 \times 10^{-12}$ and $P = 7.0 \times 10^{-9}$ respectively). **g**, RNA-seq analysis of *six6.L* (red) and *six6.S* (blue) during *X. laevis* development (left) and in adult tissues (right). Expression levels of *six6.S* were lower than those of *six6.L* at most developmental stages and in adult tissues. **h**, Diagram of *Homo sapiens*, *X. tropicalis* and *X. laevis* *six6* loci (upper panel). Magenta and black boxes indicate CNEs and exons, respectively. The phylogenetic tree analyses of *H. sapiens*, *X. tropicalis* and *X. laevis* *six6* CNEs (lower left panel) and *Six6* proteins (lower right panel). Notably, *six6.S* is more diverged from *X. tropicalis* *six6* than *six6.L*, both in the encoded protein sequences and in CNEs within 3 kb of the transcription start sites. Materials, methods and the CNE locations on genome assemblies are described in Supplementary Note 13.1. **i**, On the basis of chromatin state properties, a Random Forest machine-learning algorithm can accurately predict L versus S expression bias. The classification is based on all genes with greater than threefold expression difference at NF stage 10.5 (a set of 1,129 genes). The mean (dotted black line) of the ROC area under the curve is 0.778 (tenfold cross-validation). Features were selected using Linear Support Vector Classification and are shown in **j**. **j**, Relative importance (based on Gini impurity) of selected features used in the Random Forest classification. All features used in the classification are shown. Among various variables, the ratios of H3K4me3 and DNA methylation at the promoter contributed most to the decision tree model. A difference in p300 binding in the genomic region surrounding the gene also contributed to the Random Forest classification, as did the presence or absence of a number of specific transcription factor motifs in the promoter.

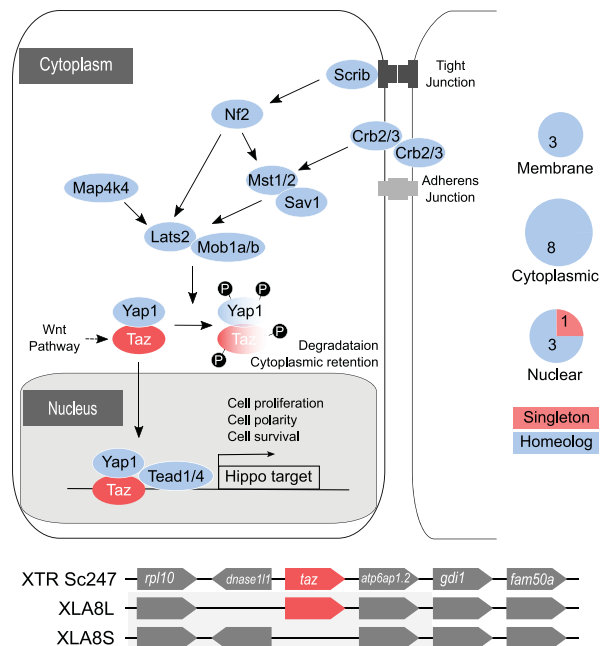
a



b

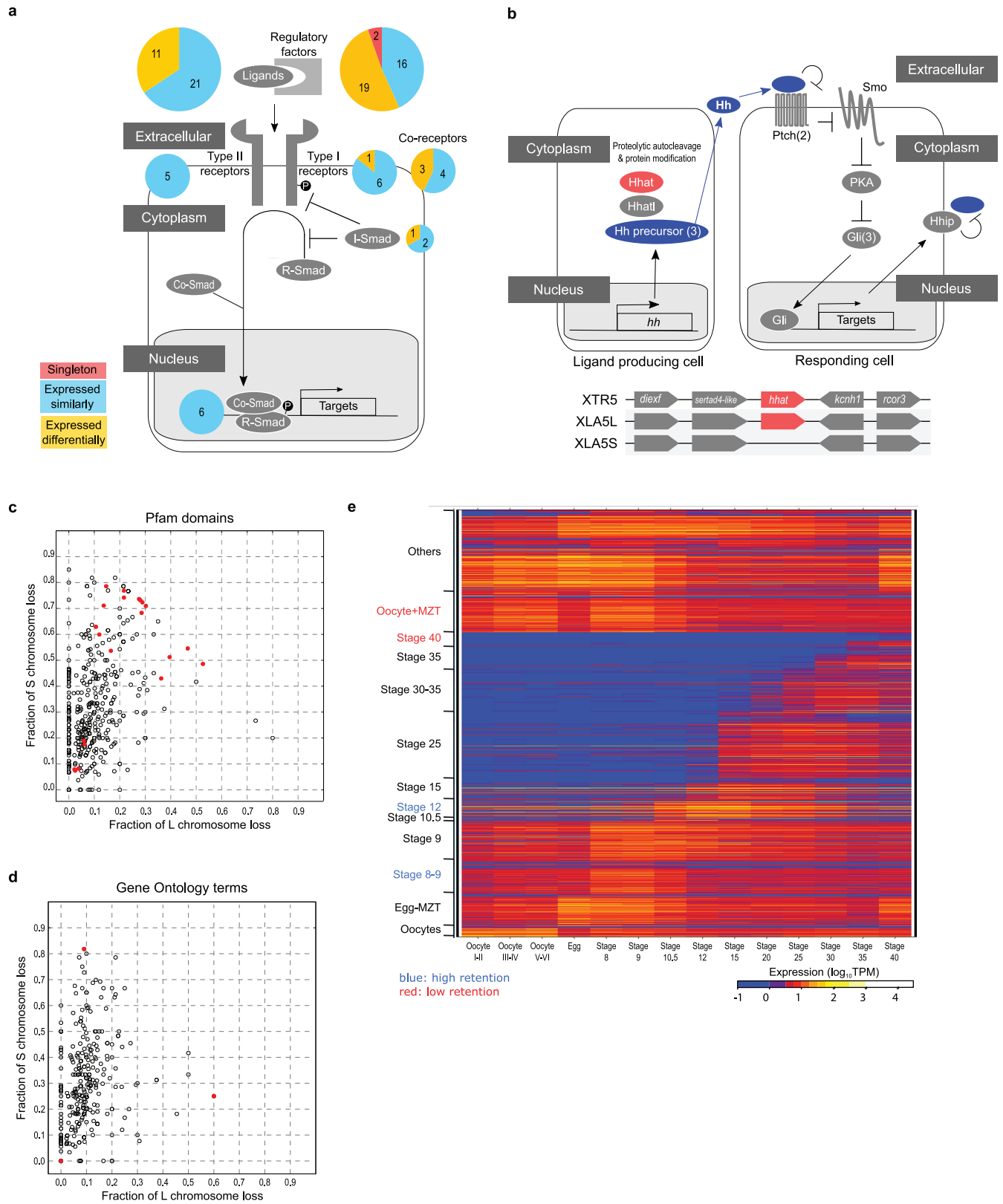


c



Extended Data Figure 9 | Examples of pathway responses. a, The Wnt pathway. Left panel, several key components of the canonical Wnt pathway in the *X. laevis* genome. The numbers in brackets show the number of paralogues. Components that have homoeologous pair of genes or singletons are shown in blue or red, respectively. Each gene (*wnt*: 21 genes, LRP: 2 genes, Fzd: 10 genes, Dvl: 3 genes, Frat(GBP): 1 gene, GSK3: 2 genes, Axin: 2 genes, βcatenin: 1 gene, APC: 2 genes, TCF/LEF: 4 genes) was classified into 4 groups according to subcellular localization, and the number of singleton and homoeologue retained genes is shown by pie charts. Right panel, synteny around four singleton genes. **b,** Cell cycle regulation. Upper right panel, diagram of the cell cycle and regulatory proteins critical to each phase. Cyclin H (*ccnh*) and Cdk7 constitute Cdk-activating kinase (CAK), a key factor required for activation of all Cdks. Genes encoding Cyclin H and Cdk7 (red), but not other regulators (blue), became singletons. Upper left panel, pie charts show the numbers of homoeologous pairs (blue) and singletons (red) in each functional category as indicated. Lower left panel, synteny of *ccnh* and

cdk7 loci in *X. tropicalis* and *X. laevis*. Lower right table, individual genes used for drawing the pie charts are shown in the table. **c,** The Hippo pathway. Upper panel, Hippo pathway components and retention of their homoeologous gene pairs. All genes for Hippo pathway components as indicated were identified in the whole genome of *X. laevis*. Blue icons indicate that both of the homoeologous genes are expressed in normal development and adult organs. The red icon, Taz, indicates a singleton. Yap is interchangeable with Taz in most cases, but TAZ, but not YAP, serves as a mediator of Wnt signalling (broken line). Pie charts show the numbers of homoeologue pairs (blue) and singletons (red) in each category of the Hippo pathway components classified according to subcellular localization. Lower panel, comparative analysis of synteny around the *taz* gene. *X. tropicalis* scaffold247 is not incorporated into the chromosome-scale assembly (v9) and hence its chromosomal location is not known yet. The p arm termini of XLA8L and XLA8S are on the left. See Supplemental Note 13 for further details.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Pathways continued. a, The TGF β pathway. Pie charts indicate the ratio of differentially expressed homoeologous pairs (orange) and singletons (red). Many of the extracellular regulatory factors are either differentially regulated or became singletons. Genes for a type I receptor, co-receptors and an inhibitory Smad are also differentially regulated. Multicopy genes such as *nodal3*, *nodal5* and *vg1* are not counted as singletons, even though those genes are deleted on S chromosomes. Instead, these and duplicated *chordin* genes are categorized into differentially regulated genes. **b,** The sonic hedgehog pathway. Upper panel, the simplified hedgehog pathway known in Shh signalling is schematically shown. Most signalling components are encoded by both homoeologous genes, whereas Hhat (shown in red) is encoded by a singleton gene. Where paralogues exist, the numbers of paralogues are shown in parentheses. In the left cell, the Shh precursor (Hh precursor) is matured through the process involving Hhat and Hhatl and secreted. In the right cell, the binding of Shh (Hh) to Ptch1 (Ptch) receptor inhibits Ptch1-mediated repression of Smo, leading to Smo activation and subsequent inhibition of PKA; otherwise PKA converts Gli activators to truncated repressors. As a consequence, Gli proteins activate target genes, such as Ptch1 and Hhip. The transmembrane protein Hhip binds Shh and suppresses Shh activity. Lower panel, schematic comparison of synteny around *hhat* genes of *X. tropicalis* chromosome 5 (top) and

X. laevis 5L chromosome (middle) and the corresponding region of *X. laevis* 5S chromosome (bottom). The diagram is not drawn to scale. **c,** Deletion rates on L (x axis) versus S (y axis) for different Pfam groups. For Pfam groups we computed the number of *X. laevis* single-copy genes (singletons) versus homoeologue pairs and computed the fraction retained. The line of expected L/S loss is based on the genome-wide average (56.4%). Red points show groups with high or low rates of loss ($P < .01$). See Supplementary Table 5 for more information. **d,** Deletion rates on L (x axis) versus S (y axis) for different stage weighted gene correlation network analysis (WGCNA)⁵⁴ groups (visualized as a heatmap in Fig. 4a). For stage WGCNA groups we computed the number of *X. laevis* single-copy genes (singletons) versus homoeologue pairs and computed the fraction retained. The line of expected L/S loss is based on the genome-wide average (56.4%). Red points show groups with high or low rates of loss ($P < .01$). **e,** Deletion rates on L (x axis) versus S (y axis) for different GO groups. For GO groups we computed the number of *X. laevis* single-copy genes (singletons) versus homoeologue pairs and computed the fraction retained. The line of expected L/S loss is based on the genome-wide average (56.4%). Red points show groups with high or low rates of loss ($P < 0.01$). See Supplementary Table 5 for more information.

Diversity-oriented synthesis yields novel multistage antimalarial inhibitors

Nobutaka Kato^{1*}, Eamon Comer^{1*}, Tomoyo Sakata-Kato², Arvind Sharma³, Manmohan Sharma³, Micah Maetani^{1,4}, Jessica Bastien¹, Nicolas M. Brancucci², Joshua A. Bittker¹, Victoria Corey⁵, David Clarke², Emily R. Derbyshire^{1,6,7}, Gillian L. Dornan⁸, Sandra Duffy⁹, Sean Eckley¹⁰, Maurice A. Itoe², Karin M. J. Koolen¹¹, Timothy A. Lewis¹, Ping S. Lui², Amanda K. Lukens^{1,2}, Emily Lund², Sandra March^{1,12}, Elamaran Meibalan², Bennett C. Meier^{1,4}, Jacob A. McPhail⁸, Branko Mitasev¹⁰, Eli L. Moss¹, Morgane Sayes¹, Yvonne Van Gessel¹⁰, Mathias J. Wawer¹, Takashi Yoshinaga¹³, Anne-Marie Zeeman¹⁴, Vicky M. Avery⁹, Sangeeta N. Bhatia^{1,12}, John E. Burke⁸, Flaminia Catteruccia², Jon C. Clardy^{1,6}, Paul A. Clemons¹, Koen J. Decherling¹¹, Jeremy R. Duvall¹, Michael A. Foley¹, Fabian Gusovsky¹⁰, Clemens H. M. Kocken¹⁴, Matthias Marti², Marshall L. Morningstar¹, Benito Munoz¹, Daniel E. Neafsey¹, Amit Sharma³, Elizabeth A. Winzeler⁵, Dyann F. Wirth^{1,2}, Christina A. Scherer¹ & Stuart L. Schreiber^{1,4}

Antimalarial drugs have thus far been chiefly derived from two sources—natural products and synthetic drug-like compounds. Here we investigate whether antimalarial agents with novel mechanisms of action could be discovered using a diverse collection of synthetic compounds that have three-dimensional features reminiscent of natural products and are underrepresented in typical screening collections. We report the identification of such compounds with both previously reported and undescribed mechanisms of action, including a series of bicyclic azetidines that inhibit a new antimalarial target, phenylalanyl-tRNA synthetase. These molecules are curative in mice at a single, low dose and show activity against all parasite life stages in multiple *in vivo* efficacy models. Our findings identify bicyclic azetidines with the potential to both cure and prevent transmission of the disease as well as protect at-risk populations with a single oral dose, highlighting the strength of diversity-oriented synthesis in revealing promising therapeutic targets.

Malaria is a deadly disease caused by protozoan parasites of the genus *Plasmodium*. Effective eradication strategies have been elusive, primarily owing to the complex life cycle of *Plasmodium* and the emergence of drug-resistant strains of *P. falciparum*, the most lethal *Plasmodium* species in humans¹. The majority of the current antimalarial drugs target the asexual blood stage of *Plasmodium*, in which they parasitize and replicate within erythrocytes². Even though liver- and transmission-stage parasites do not cause malarial symptoms, prophylaxis and transmission-blocking drugs are essential for the proactive prevention of disease epidemics and to protect vulnerable populations^{3,4}. Unfortunately, the current antimalarial drugs do not address all of the requirements for the targeting of pan-life-cycle activity. Several recent reports have described next-generation drug candidates that may achieve some of these important goals^{2,5–9}. However, eradication will require multiple innovative ways of targeting the parasite^{10–12}. The antimalarial pipeline will therefore benefit from compounds with diverse mechanisms of action, features that should help circumvent the many resistance mechanisms that render existing drugs ineffective.

We identified two key features of a successful strategy for overcoming these challenges. The first of these is the application of modern methods of asymmetric organic synthesis to create unique chemical matter; the second is to test the resulting compounds in a series of

phenotype-based screens designed to uncover agents that act on targets essential for several stages of the parasite life cycle (that is, multistage activity). We were encouraged by a small-scale pilot experiment that followed this blueprint and yielded the antimalarial agent ML238 (refs 13–15). The experiments described here excluded this earlier pilot set of compounds.

We tested synthetic compounds with structures that were inspired by the structural complexity and diversity of the entire ensemble of natural products, rather than by specific natural products. In this way, we deliberately break the link to natural selection and the limitations it provides in terms of target diversity¹⁶. A high-throughput *P. falciparum* phenotypic screen of infected erythrocytes was used to detect inhibitors of parasite growth, with counter-screens using parasites that are resistant to approved or developmental drugs, and with liver- and transmission-stage parasites used to facilitate the discovery of compounds that act through novel mechanisms of action and target multiple stages of malarial infection.

Approximately 100,000 compounds, synthesized at the Broad Institute using the build/couple/pair strategy^{17,18} of diversity-oriented synthesis (DOS), were screened against a multi-drug-resistant strain (*P. falciparum* strain Dd2) using a phenotypic blood-stage growth-inhibition assay, which models a human blood-stage infection. Compounds scored as positives were counter-screened in parallel

¹Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, Massachusetts 02142, USA. ²Harvard T.H. Chan School of Public Health, 665 Huntington Avenue Boston, Massachusetts 02115, USA. ³Molecular Medicine Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Road, New Delhi 110067, India. ⁴Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. ⁵School of Medicine, University of California, San Diego, 9500 Gilman Drive 0760, La Jolla, California 92093, USA. ⁶Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ⁷Department of Chemistry and Department of Molecular Genetics and Microbiology, Duke University, 124 Science Drive, Durham, North Carolina 27708, USA. ⁸Department of Biochemistry and Microbiology, University of Victoria, 270 Petch Hall, Victoria, British Columbia V8P 5C2, Canada. ⁹Eskitis Institute for Drug Discovery, Griffith University, Nathan Campus, Griffith University, Nathan, Brisbane, Queensland 4111, Australia. ¹⁰Eisai Inc., 4 Corporate Drive, Andover, Massachusetts 01810, USA. ¹¹TropiQ Health Sciences, Geert Grooteplein 28, Huispost 268, 6525 GA Nijmegen, The Netherlands. ¹²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 500 Main Street, Cambridge, Massachusetts 02142, USA. ¹³Eisai Co. Ltd, 5-1-3 Tokodai, Tsukuba, Ibaraki 300-2635, Japan. ¹⁴Department of Parasitology, Biochemical Primate Research Centre, 2280 GH Rijswijk, The Netherlands.

*These authors contributed equally to this work.

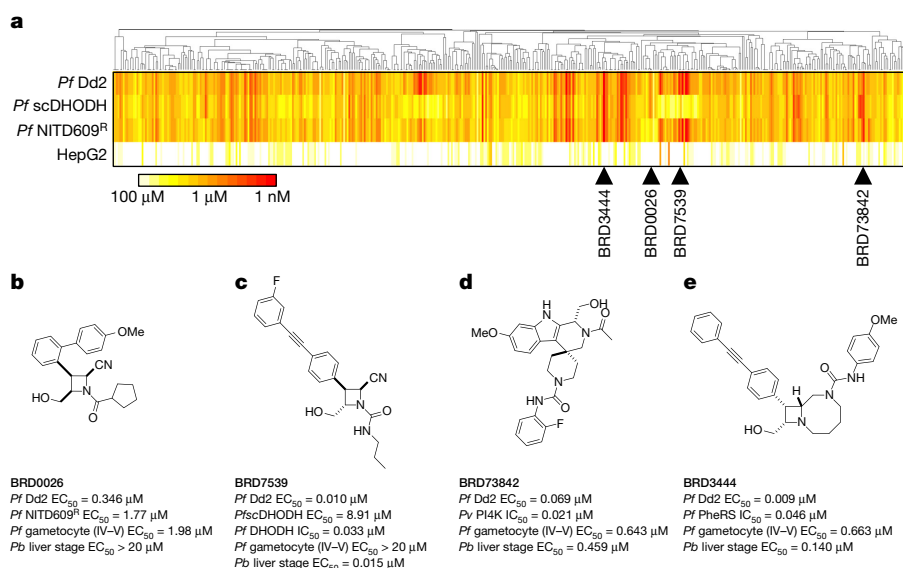


Figure 1 | Cascading triage strategy reveals targets for some of the hit compounds and highlights potential novel mechanisms of action for others. **a–e**, A total of 468 compounds ('positives' in the growth inhibition primary assay) were tested in dose against *P. falciparum* Dd2, a transgenic *P. falciparum* line expressing *Saccharomyces cerevisiae* DHODH (*Pf*scDHODH), a *P. falciparum* strain resistant to NITD609 (*Pf*NITD609^R) and a mammalian cell line (HepG2). *P. falciparum* ATPase4 is the presumed

molecular target of NITD609 (ref. 9). **a**, Compounds were clustered across the horizontal axis by structural similarity. Colours represent compound potency (EC₅₀). Two compound clusters, exemplified by BRD0026 (**b**) and BRD7539 (**c**), showed selectively reduced potency against the *Pf* NITD609^R and *Pf*scDHODH strains, respectively, while BRD73842 (**d**) and BRD3444 (**e**) were equipotent across the three *P. falciparum* strains. *Pb*, *P. berghei*; *Pf*, *P. falciparum*; *Pv*, *P. vivax*; PheRS, phenylalanyl-tRNA synthetase.

against a panel of parasite isolates and diverse drug-resistant clones to deprioritize compounds with previously identified mechanisms of action (Fig. 1a and Supplementary Tables 1, 2). After evaluating results from assays against the liver-stage (*Plasmodium berghei* strain ANKA) and transmission-stage (*P. falciparum* strain 3D7) parasites, four chemical series with additional liver-stage and/or transmission-blocking activities (BRD0026, BRD7539, BRD73842 and BRD3444; Fig. 1b–e, Extended Data Table 1 and Supplementary Tables 1, 2) were selected. This layered screening process also yielded other series not described here that may merit attention in the future (available at the Malaria Therapeutics Response Portal, <http://portals.broadinstitute.org/mtrp/>). Underlying features of DOS helped to guide the selection and development of the four nominated series. The compound collection includes stereoisomeric families that yield stereochemistry-based structure–activity relationships (SSAR); their inclusion indicated the possibility of selective interactions with targets. The short, modular pathways, entailing inter- and intramolecular coupling reactions, facilitate medicinal chemistry optimization. Three of the four series yielded new compound scaffolds against known targets. These include: (i) disruptors of sodium ion regulation mediated by *P. falciparum* ATPase4 (ref. 9; BRD0026 is active against asexual and late sexual blood stages of parasites, Fig. 1b and Extended Data Fig. 1a–d); (ii) potent and selective inhibitors of *P. falciparum* dihydroorotate dehydrogenase (*pf* DHODH)¹⁹ (BRD7539 is active against liver-stage and asexual blood-stage parasites; Fig. 1c and Extended Data Fig. 1e–h); and (iii) potent and selective inhibitors of *P. falciparum* phosphatidylinositol-4-kinase (*pf* PI4K)^{20,21} (BRD73842 is active against liver-stage, asexual and late sexual blood-stage parasites; Fig. 1d, Extended Data Figs 1i–m, 2a and Supplementary Table 3). The fourth series was found to inhibit a previously unknown antimalarial target and is characterized in detail below.

Bicyclic azetidines inhibit cytosolic *Pf* PheRS

The bicyclic azetidine BRD3444 showed multistage activity *in vitro* (*P. falciparum* Dd2, blood stage, half-maximal effective concentration (EC₅₀) = 9 nM; *P. falciparum* 3D7, transmission stage, gametocyte IV–V, EC₅₀ = 663 nM; *P. berghei* strain ANKA, liver stage, EC₅₀ = 140 nM; Fig. 1e, Extended Data Table 1 and Supplementary Table 1). To elucidate the mechanism of action of the bicyclic

azetidine series, three resistant lines were evolved against BRD1095 (Fig. 2a and Extended Data Fig. 2b), a derivative of BRD3444 with increased aqueous solubility, from eight independent cultures (> 8 × 10⁹ inocula). After more than 3 months of drug pressure, EC₅₀ values were increased by 4–84-fold. Two clones were obtained from each culture and genomic DNA from each clone was analysed via whole-genome sequencing (Fig. 3a, b and Supplementary Table 4). Analysis of resistant clones revealed that each had at least one non-synonymous single-nucleotide variant (SNV) in the PF3D7_0109800 locus, which is predicted to encode the alpha subunit of the cytosolic phenylalanyl-tRNA synthetase (*Pf* PheRS)

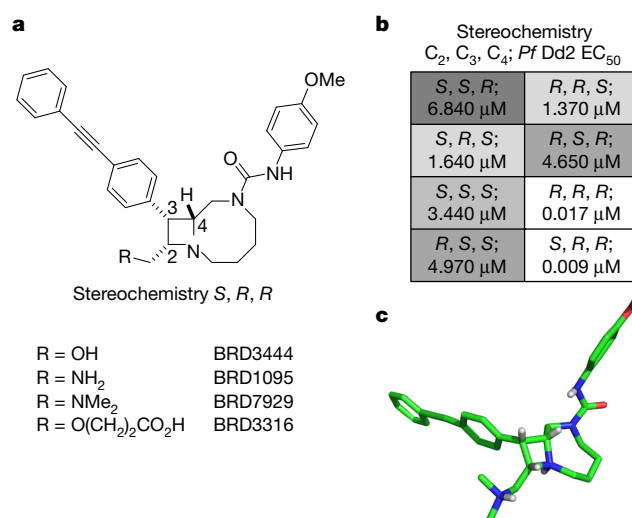


Figure 2 | Structures of key compounds, SSAR study of BRD3444 and X-ray crystal structure of BRD7929. **a**, Structures of four bicyclic azetidine compounds. **b**, SSAR of BRD3444 showing that stereoisomers at the C₂ position are equipotent, which suggests that this position is not necessary for activity. **c**, X-ray crystal structure of BRD7929 showing 3D conformation (BRD7929 was crystallized as a salt with two equivalents of L-tartaric acid; only the structure of BRD7929 is shown for clarity).

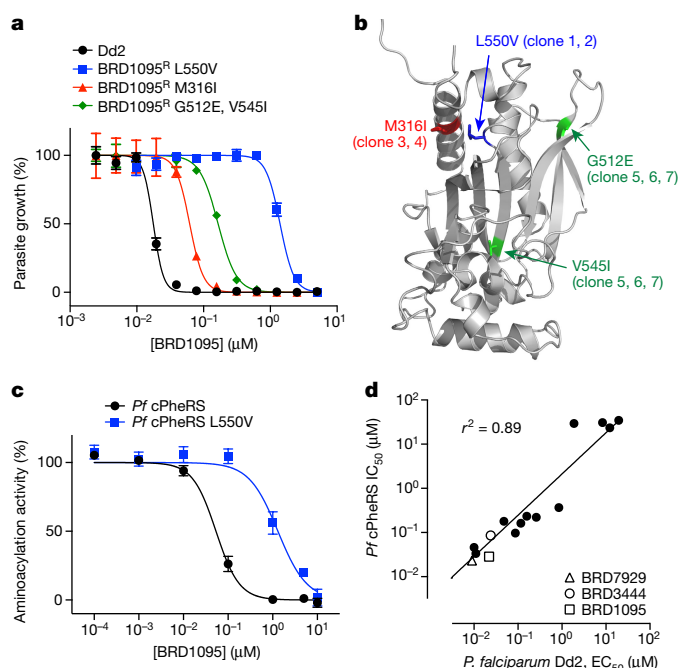


Figure 3 | The bicyclic azetidine series targets the cytoplasmic *Pf* PheRS. **a**, *P. falciparum* Dd2 clones resistant to BRD1095, a derivative of BRD3444 with increased aqueous solubility, were selected *in vitro* and non-synonymous SNVs were identified via whole-genome sequencing. All clones from three individual flasks contained non-synonymous SNVs within the PF3D7_0109800 locus, which encodes the alpha subunit of the cytoplasmic PheRS. **b**, The non-synonymous SNVs identified in clones from flask 1 (red), flask 2 (blue), and flask 3 (green) are shown overlaid on a homology model based on the human cytoplasmic PheRS (PDB accession 3LAG) generated in PyMol. **c**, BRD1095 was assayed against purified recombinant proteins of wild-type cytosolic *Pf* PheRS and a mutant containing a SNV (giving a L550V substitution), identified from the resistant strain. IC_{50} value of the wild-type PheRS was $0.045 \mu\text{M}$, whereas the IC_{50} value for BRD1095^{L550V} was $1.30 \mu\text{M}$ (data are mean \pm s.d. for two biological and two technical replicates). **d**, The bicyclic azetidine series showed a strong correlation between blood-stage growth inhibition and biochemical inhibition of cytosolic *Pf* PheRS activity. We assayed 15 bicyclic azetidine analogues with varying potency against blood-stage parasites (Dd2 strain) against purified recombinant *Pf* PheRS. The biochemically derived IC_{50} values correlate strongly ($r^2 = 0.89$) with the EC_{50} values determined using the blood-stage growth inhibition assay (see Extended Data Table 2 for structure–activity relationship study and chemical structures).

of *P. falciparum* (ref. 22). Examination of more than 100 drug-resistant *P. falciparum* clones failed to reveal even a single SNV in the PF3D7_0109800 locus, indicating that the probability of *Pf* PheRS having three independent mutations by chance is very low. To confirm that cytosolic PheRS is the molecular target of BRD1095, the compound was assayed against purified recombinant proteins. BRD1095 inhibited the aminoacylation activity of recombinant *Pf* PheRS in a concentration-dependent manner (half-maximal inhibitory concentration (IC_{50}) = 46 nM ; Fig. 3c). We also reasoned that if the primary antiparasitic mechanism of the bicyclic azetidine series was via inhibition of *Pf* PheRS activity, then IC_{50} values for the aminoacylation activity of purified recombinant *Pf* PheRS proteins should correlate with EC_{50} values obtained in parasite growth inhibition assays. Indeed, a high correlation between the two parameters ($r^2 = 0.89$) was observed using 16 synthetic analogues of BRD1095 covering a range of activities (Fig. 3d and Extended Data Table 2). This notable correlation, together with the aforementioned genetic evidence, indicates that cytosolic *Pf* PheRS is the relevant molecular target of the bicyclic azetidine series. In addition, supplementation with exogenous L-phenylalanine (but not D-phenylalanine, L-aspartic acid, L-threonine or L-tyrosine) to the

in vitro culture medium increased the EC_{50} value of BRD1095 in a concentration-dependent manner (Supplementary Table 5).

Owing to its newfound susceptibility to inhibition, *Pf* PheRS joins the aminoacyl-tRNA synthetase class of emerging targets for antimalarial agents^{23–29}. Although they share common tRNA esterification catalytic activities, these proteins are structurally diverse and physiologically distinct enzymes. The target described here (*P. falciparum* cytosolic PheRS) is unique as it is the first member of the class in which inhibition, as we will describe, results in elimination of asexual blood-, liver- and transmission-stage parasites, preventing disease transmission, ensuring prophylaxis and providing single-dose cures of the disease in mouse models of malaria.

Optimization of the bicyclic azetidine series

BRD3444 exhibited poor solubility ($<1 \mu\text{M}$ in PBS), high intrinsic clearance in human and mouse microsomes ($\text{Cl}_{\text{int}} = 142$ and $248 \mu\text{L min}^{-1} \text{mg}^{-1}$, respectively) and a high volume of distribution ($V_{\text{ss}} = 12 \text{ l kg}^{-1}$; all data found in Extended Data Table 3). These results translated to a half-life of 3.7 h in an intravenous pharmacokinetic study in CD-1 mice. Analysis of all eight stereoisomers of BRD3444 included in the primary screen revealed that activity against *P. falciparum* Dd2 parasites was predominantly found among two isomers differing in stereochemistry at the C_2 position (Fig. 2a, b). Therefore, we postulated that the C_2 position could be manipulated without loss of *in vitro* potency and could be used to improve the physicochemical and pharmacokinetic properties of the series. The modular synthetic pathway facilitated the synthesis of advanced analogues that included BRD1095 and BRD7929, in which the hydroxymethyl group at position C_2 is replaced with aminomethyl and dimethylaminomethyl substituents, respectively. These bicyclic azetidines showed improved solubility (25 and $15 \mu\text{M}$ in PBS, respectively) and greatly improved intrinsic clearance in mouse microsomes (<20 and $21 \mu\text{L min}^{-1} \text{mg}^{-1}$, respectively), while retaining *in vitro* potency. In an intravenous and oral pharmacokinetic study in mice, both BRD1095 and BRD7929 displayed greatly improved blood clearance relative to BRD3444. BRD7929 also displayed good bioavailability (80%), superior to that of BRD1095 (50%), and improved *in vitro* potency against *P. cynomolgi* and *P. falciparum* liver-stage and *P. falciparum* transmission-stage parasites (Extended Data Table 1). BRD7929 showed a high V_{ss} of 241 l kg^{-1} (Extended Data Table 3), which, together with a low blood clearance, translated to a long half-life (32 h), making this compound suitable for single-dose oral treatments. The synthesis pathway enabled the laboratory preparation of 7.5 g of BRD7929 for further testing.

BRD7929 shows *in vivo* efficacy against all life stages

We evaluated the multistage activity of BRD7929 using mouse malaria models. When BRD7929 activity was evaluated in the blood-stage model with the rodent malaria parasite *P. berghei* using a luciferase reporter, all infected CD-1 mice treated with a single oral 25 mg kg^{-1} or 50 mg kg^{-1} dose became parasite-free and remained so up to the 30-day end-point based on bioluminescent imaging (Extended Data Fig. 3a, b). To evaluate the therapeutic potential of this series, the *in vivo* efficacy of BRD7929 against the human malaria parasite *P. falciparum* was determined. Approximately 48 h after inoculation with the blood-stage *P. falciparum* 3D7^{HLH/BRD} (expressing firefly luciferase), non-obese diabetic/severe combined immunodeficiency (NOD/SCID) *Il2r $\gamma^{-/-}$* mice engrafted with human erythrocytes (huRBC NSG) were treated with a single dose of BRD7929 and monitored for 30 days (Fig. 4a and Extended Data Fig. 3c). At 25 mg kg^{-1} (area under curve (AUC) = $62.8 \mu\text{M h}$) and 50 mg kg^{-1} (AUC = $125.6 \mu\text{M h}$), a rapid decrease in parasite-associated bioluminescence was observed, while at 6.25 mg kg^{-1} (AUC = $15.7 \mu\text{M h}$) the rate of the loss of bioluminescence was slower. All huRBC NSG mice treated with single oral 12.5 mg kg^{-1} (AUC = $31.4 \mu\text{M h}$), 25 mg kg^{-1} or 50 mg kg^{-1} doses were parasite-free for 30 days based on bioluminescent imaging. The AUC in

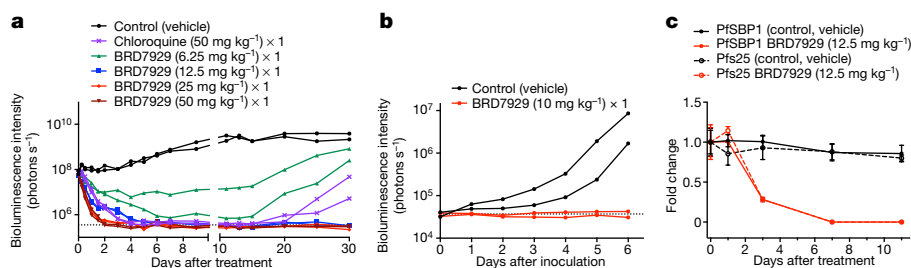


Figure 4 | In vivo efficacy studies of BRD7929 using *P. falciparum* and humanized mouse models. **a**, huRBC NSG mice were inoculated with *P. falciparum* (3D7^{HLH/BRD}) blood-stage parasites 48 h before treatment and BRD7929 was administered as a single 50, 25, 12.5 or 6.25 mg kg⁻¹ oral dose at 0 h ($n = 2$ for each group, this study was conducted once). Infections were monitored using the *in vivo* imaging system (IVIS). Bioluminescent intensity was quantified from each mouse and plotted against time. The dotted horizontal line represents the mean bioluminescence intensity level obtained from all the animals before the parasite inoculation. No recrudescence was observed as low as a single 25 mg kg⁻¹ dose of BRD7929 in the infected animals (see Extended Data Fig. 3b). **b**, huHep FRG-knockout mice were inoculated intravenously with *P. falciparum* (NF54HT-GFP-luc) sporozoites. BRD7929 was administered as a single 10 mg kg⁻¹ oral dose 1 day after inoculation, and daily engraftment of human erythrocytes was initiated 5 days after inoculation ($n = 2$ for each group, this study was conducted once).

the 25 mg kg⁻¹ single-dose cure observed in the model with *P. berghei* is estimated to be 27.5 μMh based on pharmacokinetic studies with CD-1 mice. Thus, single-dose cures were observed in the *P. berghei* CD-1 and *P. falciparum* huRBC NSG mouse models at similar drug exposure levels (AUC = 27.5 and 31.4 μMh, respectively), suggesting that the efficacy against the two *Plasmodium* species is comparable.

In a *P. berghei* liver-stage model, none of the CD-1 mice that were treated with a single dose of 5 or 25 mg kg⁻¹ BRD7929 developed blood-stage parasitaemia within a 30-day period following *P. berghei* sporozoite inoculation (Extended Data Fig. 4a, b). Furthermore, mice were treated with a single dose of 10 mg kg⁻¹ BRD7929 at various time points before sporozoite inoculation and during liver-stage infection (Extended Data Fig. 4c). All mice treated within the 3 days before inoculation and during liver-stage infection were completely free of blood-stage parasites for the duration of the experiment (32 days), indicating that BRD7929 has potent causal prophylaxis activity. Next, 1 day after inoculation with *P. falciparum* (NF54HT-GFP-luc)³⁰ sporozoites, FRG knockout (*Fah*^{-/-} *Rag2*^{-/-} *Il2rg*^{-/-}, heavily immunosuppressed) C57BL/6 mice transplanted with human hepatocytes (huHep FRG knockout)³¹ were treated with a single oral dose of BRD7929 (10 mg kg⁻¹). Human erythrocytes were intraperitoneally injected daily from 5 to 7 days after inoculation. A gradual increase was detected in parasite liver-stage-associated bioluminescence signals from the lower pectoral and upper abdominal regions of the control (vehicle-treated) mice, whereas no increase in bioluminescence signals was observed from the BRD7929-treated mice (Fig. 4b and Extended Data Fig. 5a). For quantitative reverse transcription PCR (qRT-PCR) analysis³², blood samples were also collected 7 days after inoculation (the first day of the blood stage)³¹ and evaluated for the presence of the blood-stage transcript PF3D7_1120200 (expressing the *P. falciparum* ubiquitin-conjugating enzyme, UCE) (Extended Data Fig. 5b). The presence of the blood-stage marker was not detected in samples from the BRD7929-treated mice, indicating that BRD7929 eliminated the liver-stage parasites.

Finally, to examine whether BRD7929 has activity against mature gametocytes and prevents parasite transmission to mosquitoes *in vivo*, CD-1 mice infected with *P. berghei* were treated with a single oral dose of BRD7929 2 days before exposure to female *Anopheles stephensi* mosquitoes. One week later, the midguts of the blood-fed mosquitoes were dissected and the number of oocysts was counted (Extended Data Fig. 6a–c). No oocysts were detected in

infections were monitored using IVIS. The dotted horizontal line represents the mean bioluminescence intensity level obtained from all the animals before the sporozoite inoculation. No increase in bioluminescence intensity level was observed from the mice treated with BRD7929 (see Extended Data Fig. 5a). **c**, huRBC NSG mice were infected with blood-stage *P. falciparum* (3D7^{HLH/BRD}) parasites for 2 weeks (allowing the gametocytes to mature fully) and were treated with a single oral dose of BRD7929 (12.5 mg kg⁻¹). Blood samples were collected for 11 days and analysed for the presence of the asexual marker SBP1 and the mature gametocyte marker Pfs25 using qRT-PCR ($n = 2$ for each group, this study was conducted once). The transcription of both SBP1 and Pfs25 decreased to undetectable levels 7 days after treatment, strongly suggesting that BRD7929 eliminates both asexual and gametocyte stages and is capable of preventing parasite transmission to the mosquito (data are mean ± s.d. for three technical replicates for each biological sample).

midguts dissected from mosquitoes fed on mice treated with 5 or 20 mg kg⁻¹ BRD7929, concentrations below those found to be efficacious against asexual blood-stage parasites. To determine whether BRD7929 showed *in vivo* efficacy against *P. falciparum* in humanized mouse models, huRBC NSG mice were infected with blood-stage *P. falciparum* 3D7^{HLH/BRD} parasites for 2 weeks to allow the development of mature gametocytes. Subsequently, mice were treated with a single oral dose of BRD7929 (12.5 mg kg⁻¹, AUC = 31.4 μMh). Blood samples were collected for 11 days after treatment and analysed for the presence of the late-sexual-stage-specific transcript of *Pfs25* (expressing *P. falciparum* 25 kDa ookinete surface-antigen precursor, PF3D7_1031000) using qRT-PCR³² (Fig. 4c and Extended Data Fig. 6d–f). The transcription of *Pfs25* decreased to undetectable levels 7 days after treatment. Previous literature reports of *in vitro* cellular sensitivity showed that the *Pfs25* marker had a detection limit of 0.02–0.05 gametocytes μl⁻¹ (ref. 33), strongly suggesting that BRD7929 has late-stage gametocidal activity and is capable of preventing the transmission of parasites to the mosquito vector at the same level of exposure as that achieves a single-dose cure in the blood stage.

Safety optimization of the bicyclic azetidine series

While no significant cytotoxicity was observed with BRD3444 and BRD3316, moderate cytotoxicity was observed for bicyclic azetidines BRD7929 (half-maximal cytotoxic concentration (CC₅₀) = 9 μM) and BRD1095 (CC₅₀ = 16 μM) in the HepG2 cell line (Extended Data Fig. 7a). Both BRD1095 and BRD7929 showed inhibition of *I_{Kr}* (encoded by *KCNH2*, also known as *hERG*) (IC₅₀ = 5.1 and 2.1 μM, respectively; Extended Data Table 3). Medicinal chemistry efforts have shown that mitigation of ion-channel toxicity is possible while maintaining biological activity; for example, BRD3316 shows no significant inhibition of *I_{Kr}* at >10 μM, indicating that cardiotoxicity is not intrinsically linked to this series. While BRD3444 showed time-dependent inhibition of CYP3A4, BRD7929 showed no inhibition of any of the major human cytochrome P450 (CYP) isoforms (Extended Data Fig. 7a). No phototoxicity was observed with this series in BALB/c 3T3 mouse fibroblasts following exposure to UVA light. BRD7929 and BRD3316 show desirable pharmacokinetic properties, including good oral bioavailability ($F = 80$ and 63%, respectively). In addition, BRD7929 has a long half-life that enables single-dose treatment. Based on *in vitro* microsomal stability data, BRD7929 and advanced analogues in this series are likely to have a similar profile in

humans, as metabolic clearance was low for both mouse and human species (Extended Data Table 3). BRD7929 was determined to be non-mutagenic using an Ames test in the presence or absence of S9 mix using the *Salmonella typhimurium* strains TA100, TA1535, TA98, TA1537 and *Escherichia coli* strain WP2uvrA (Supplementary Table 6). Histopathological analysis of mice treated at a high dose (100 mg kg^{-1} , estimated C_{max} and AUC are $5.4 \mu\text{M}$ and $110 \mu\text{M h}$, respectively) showed no adverse findings in the limited number of organs examined (Extended Data Fig. 7b). Additional studies involving a wider range of organs, doses and compounds will be needed to assess the toxicity of these and related compounds more thoroughly. In NSG mice the estimated C_{max} and AUC of the single-dose cure are 833 nM and $31.4 \mu\text{M h}$, respectively, affording a 6.5-fold safety margin with respect to C_{max} .

Although the emergence of resistance *in vitro* does not necessarily imply that it will happen *in vivo*, it is indicative of any mechanisms of resistance that could arise in the future. To examine the propensity of *de novo* resistance selection, *P. falciparum* Dd2 cultures with initial inocula ranging from 10^5 to 10^9 parasites were maintained in medium supplemented with 20 nM BRD7929 (the EC_{90} of strain Dd2) and monitored for 60 days to identify recrudescence parasitaemia (Extended Data Fig. 7c, d). No recrudescence was observed in Dd2 cultures exposed to a constant pressure of BRD7929, whereas the minimum inoculum of resistance for atovaquone ($\text{EC}_{90} = 2 \text{ nM}$) was 10^7 , consistent with previous reports³⁴.

Discussion

Malaria remains one of the deadliest infectious diseases. Available therapeutic agents are already limited in their efficacy, and drug resistance threatens to diminish our ability to prevent and treat the disease further. Despite a renewed effort to identify compounds with antimalarial activity, the drug discovery and development pipeline lacks target diversity and most malaria drugs are only efficacious during the asexual blood stage of parasite infection.

In these studies, we attempted to identify new antimalarial targets by screening a diverse collection of 100,000 compounds with three-dimensional topographic features derived from stereochemical and skeletal elements that are common in natural products but underrepresented in typical screening collections—compounds now accessible using DOS. The compounds are formed in short, modular syntheses that facilitate chemical optimization and manufacturing^{35,36} and have computed physical properties aimed at accelerating drug discovery³⁷. We used a primary phenotypic screen to identify a subset of compounds that inhibits parasite growth, counter-screens to prioritize molecules with both novel mechanisms of action and activity at multiple stages of the parasite life cycle, and genetic and biochemical studies to illuminate mechanisms of action. These efforts yielded several series of multiple-stage antimalarial compounds with unique scaffolds that modulate both recently described and established molecular targets.

An earlier pilot study tested key elements of the process above using a distinct 8,000-member DOS library, leading to the discovery of ML238 (refs 13, 14), a molecule that inhibits parasite growth with nanomolar potency by targeting the reductase domain of *P. falciparum* cytochrome *b* (the Q_i site), in contrast to the antimalarial agent atovaquone, which targets the oxidase domain of *P. falciparum* cytochrome *b* (the Q_o site). The study presented here led to many candidate antimalarial agents. We have, thus far, characterized four of these compound series, namely BRD0026 (targeting *P. falciparum* ATPase4), BRD7539 (targeting *P. falciparum* DHODH), BRD73842 (targeting *P. falciparum* PI4K) and BRD3444 (targeting *P. falciparum* cytoplasmic PheRS). These series were prioritized as they showed *in vitro* activity against multiple stages of the *P. falciparum* life cycle, and this was subsequently confirmed *in vivo*. We anticipate that additional compound series uncovered by these experiments, made available via the Malaria Therapeutics Response Portal (<http://portals.broadinstitute.org/mtrp/>), will target additional proteins that function as multiple-stage vulnerabilities in *Plasmodium* and other Apicomplexa pathogens.

Until now, natural products and synthetic drug-like compounds have served as the primary sources of antimalarial drugs. As parasitic susceptibility to traditional chemotypes decreases, it is becoming increasingly necessary to discover lead compounds that are unaffected by existing mechanisms of resistance. DOS coupled with phenotypic screening offers a systematic means to address this need. The results reported here describe a new target and chemotype—*Pf* PheRS and bicyclic azetidines such as BRD3316 and BRD7929—that have demonstrated the lowest-concentration single-dose cure of three promising next-generation antimalarials in the pipeline^{9,38,39} using two mouse models. Single-dose treatments facilitate compliance and overcome cost challenges in resource-deficient regions⁴⁰. The ability of BRD7929 to eliminate blood-stage (both asexual and sexual) and liver-stage parasites suggests bicyclic azetidines have the potential to cure the disease, provide prophylaxis and prevent disease transmission.

Our findings suggest that DOS-derived compound collections, which comprise three-dimensional structures reminiscent of natural products that have yielded many small-molecule probes of diverse mammalian processes^{41,42}, are also a rich resource for identifying targets and readily optimized chemical scaffolds to supplement the current antimalarial pipeline.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 October 2015; accepted 31 August 2016.

Published online 7 September 2016.

- Wells, T. N. Discovering and developing new medicines for malaria control and elimination. *Infect. Disord. Drug Targets* **13**, 292–302 (2013).
- Flannery, E. L., Chatterjee, A. K. & Winzeler, E. A. Antimalarial drug discovery—approaches and progress towards new medicines. *Nat. Rev. Microbiol.* **11**, 849–862 (2013).
- Ariey, F. et al. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50–55 (2014).
- Campo, B., Vandal, O., Wesche, D. L. & Burrows, J. N. Killing the hypnozoite—drug discovery approaches to prevent relapse in *Plasmodium vivax*. *Pathog. Glob. Health* **109**, 107–122 (2015).
- Hameed P, S. et al. Triaminopyrimidine is a fast-killing and long-acting antimalarial clinical candidate. *Nat. Commun.* **6**, 6715 (2015).
- Jiménez-Díaz, M. B. et al. Improved murine model of malaria using *Plasmodium falciparum* competent strains and non-myelodepleted NOD-scid IL2R γ^{null} mice engrafted with human erythrocytes. *Antimicrob. Agents Chemother.* **53**, 4533–4536 (2009).
- Phillips, M. A. et al. A long-duration dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria. *Sci. Transl. Med.* **7**, 296ra111 (2015).
- Roberts, L. & Enserink, M. Malaria. Did they really say ... eradication? *Science* **318**, 1544–1545 (2007).
- Rottmann, M. et al. Spiroindolones, a potent compound class for the treatment of malaria. *Science* **329**, 1175–1180 (2010).
- Gamo, F.-J. et al. Thousands of chemical starting points for antimalarial lead identification. *Nature* **465**, 305–310 (2010).
- Guiguemde, W. A. et al. Chemical genetics of *Plasmodium falciparum*. *Nature* **465**, 311–315 (2010).
- Meister, S. et al. Imaging of *Plasmodium* liver stages to drive next-generation antimalarial drug discovery. *Science* **334**, 1372–1377 (2011).
- Comer, E. et al. Diversity-oriented synthesis-facilitated medicinal chemistry: toward the development of novel antimalarial agents. *J. Med. Chem.* **57**, 8496–8502 (2014).
- Heidebrecht, R. W. Jr et al. Diversity-oriented synthesis yields a novel lead for the treatment of malaria. *ACS Med. Chem. Lett.* **3**, 112–117 (2012).
- Lukens, A. K. et al. Diversity-oriented synthesis probe targets *Plasmodium falciparum* cytochrome *b* ubiquinone reduction site and synergizes with oxidation site inhibitors. *J. Infect. Dis.* **211**, 1097–1103 (2015).
- Dancik, V., Seiler, K. P., Young, D. W., Schreiber, S. L. & Clemons, P. A. Distinct biological network properties between the targets of natural products and disease genes. *J. Am. Chem. Soc.* **132**, 9259–9261 (2010).
- Burke, M. D. & Schreiber, S. L. A planning strategy for diversity-oriented synthesis. *Angew. Chem. Int. Edn Engl.* **43**, 46–58 (2004).
- Marcaurelle, L. A. et al. An aldol-based build/couple/pair strategy for the synthesis of medium- and large-sized rings: discovery of macrocyclic histone deacetylase inhibitors. *J. Am. Chem. Soc.* **132**, 16962–16976 (2010).
- Painter, H. J., Morrissey, J. M., Mather, M. W. & Vaidya, A. B. Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. *Nature* **446**, 88–91 (2007).
- McNamara, C. W. et al. Targeting *Plasmodium* PI(4)K to eliminate malaria. *Nature* **504**, 248–253 (2013).

21. Ghidelli-Disse, S. *et al.* Identification of *Plasmodium* PI4 kinase as target of MMV390048 by chemoproteomics. *Malar. J.* **13** (suppl. 1), 38 (2014).
22. Sharma, A. & Sharma, A. *Plasmodium falciparum* mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. *Biochem. J.* **465**, 459–469 (2015).
23. Pham, J. S. *et al.* Aminoacyl-tRNA synthetases as drug targets in eukaryotic parasites. *Int. J. Parasitol.* **4**, 1–13 (2014).
24. Herman, J. D. *et al.* The cytoplasmic prolyl-tRNA synthetase of the malaria parasite is a dual-stage target of febrifugine and its analogs. *Sci. Transl. Med.* **7**, 288ra77 (2015).
25. Hussain, T., Yogavel, M. & Sharma, A. Inhibition of protein synthesis and malaria parasite development by drug targeting of methionyl-tRNA synthetases. *Antimicrob. Agents Chemother.* **59**, 1856–1867 (2015).
26. Novoa, E. M. *et al.* Analogs of natural aminoacyl-tRNA synthetase inhibitors clear malaria *in vivo*. *Proc. Natl Acad. Sci. USA* **111**, E5508–E5517 (2014).
27. Hoepfner, D. *et al.* Selective and specific inhibition of the *Plasmodium falciparum* lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. *Cell Host Microbe* **11**, 654–663 (2012).
28. Istvan, E. S. *et al.* Validation of isoleucine utilization targets in *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **108**, 1627–1632 (2011).
29. Bhatt, T. K. *et al.* A genomic glimpse of aminoacyl-tRNA synthetases in malaria parasite *Plasmodium falciparum*. *BMC Genomics* **10**, 644 (2009).
30. Vaughan, A. M. *et al.* A transgenic *Plasmodium falciparum* NF54 strain that expresses GFP-luciferase throughout the parasite life cycle. *Mol. Biochem. Parasitol.* **186**, 143–147 (2012).
31. Vaughan, A. M. *et al.* *Plasmodium falciparum* genetic crosses in a humanized mouse model. *Nat. Methods* **12**, 631–633 (2015).
32. Chang, H.-H. *et al.* Persistence of *Plasmodium falciparum* parasitemia after artemisinin combination therapy: evidence from a randomized trial in Uganda. *Sci. Rep.* **6**, 26330 (2016).
33. Joice, R. *et al.* *Plasmodium falciparum* transmission stages accumulate in the human bone marrow. *Sci. Transl. Med.* **6**, 244re5 (2014).
34. Ding, X. C., Ubben, D. & Wells, T. N. A framework for assessing the risk of resistance for anti-malarials in development. *Malar. J.* **11**, 292 (2012).
35. Dandapani, S. & Marcaurelle, L. A. Grand challenge commentary: Accessing new chemical space for ‘undruggable’ targets. *Nat. Chem. Biol.* **6**, 861–863 (2010).
36. Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).
37. Lowe, J. T. *et al.* Synthesis and profiling of a diverse collection of azetidine-based scaffolds for the development of CNS-focused lead-like libraries. *J. Org. Chem.* **77**, 7187–7211 (2012).
38. Baragaña, B. *et al.* A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature* **522**, 315–320 (2015).
39. Younis, Y. *et al.* 3,5-Diaryl-2-aminopyridines as a novel class of orally active antimalarials demonstrating single dose cure in mice and clinical candidate potential. *J. Med. Chem.* **55**, 3479–3487 (2012).
40. Burrows, J. N., van Huijsduijnen, R. H., Möhrle, J. J., Oeuvray, C. & Wells, T. N. C. Designing the next generation of medicines for malaria control and eradication. *Malar. J.* **12**, 187 (2013).
41. Ng, P. Y., Tang, Y., Knosp, W. M., Stadler, H. S. & Shaw, J. T. Synthesis of diverse lactam carboxamides leading to the discovery of a new transcription-factor inhibitor. *Angew. Chem. Int. Edn Engl.* **46**, 5352–5355 (2007).
42. Yu, C. *et al.* High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by the Bill and Melinda Gates Foundation (grant OPP1032518 to S.L.S., grant OPP1054480 to E.A.W. and D.F.W., grant OPP1023607 to S.N.B.), the Global Health Innovative Technology Fund (grant G2014-107 to S.L.S.), Medicines for Malaria Venture and the Wellcome Trust (grant WT078285 to C.H.M.K.), a New Investigator and Open Operating Grant from Canadian Institute of Health Research (grant FRN 142393 to J.E.B.) and Medicines for Malaria Venture (grant 12-2400 to V.M.A.). S.L.S. is an Investigator at the Howard Hughes Medical Institute. Mi.M. was supported by a fellowship from the National Science Foundation (DGE1144152). The authors thank R. Elliott, K. Duncan, and O. Vandal as well as J. Burrows, J. Duffy, F. Escudié and colleagues for discussions and access to invaluable scientific and experimental resources; K. Emmith for assistance with data processing and management; I. Goldowitz for assistance with establishing a gametocyte assay; N. van der Werff for technical assistance with the *P. cynomolgi* assay; J. Kotz and B. Meillo for discussions and assistance with the manuscript; J. Pu, M. Leighty, B. Braibant, S. LeQuement and J. Beaudoin for assistance with compound synthesis; E. Garcia-Rivera for assistance with molecular modelling; A. Hakura for performing the Ames test; Broad Institute Comparative Medicine Platform and Facility for assistance with animal studies; and the Broad Institute Compound Management and analytical teams for assistance with compound access and characterization. We also acknowledge WuXi AppTec and ChemPartner Co., Ltd for *in vitro* and *in vivo* pharmacokinetics assays, and Cyprotek for the phototoxicity analysis. *P. falciparum* scDHODH transgenic strain was a gift from A. B. Vaidya, *P. falciparum* 3D7^{HLH} strain from D. Fidock and *P. falciparum* NF54HT-GFP-luc from S. H. Kappe.

Author Contributions The author contributions are detailed in the Supplementary Information.

Author Information The whole-genome sequencing analysis data in this study have been deposited to the Short Read Archive; accession code SRP064988 and the x-ray crystal structure data for BRD7929 have been deposited to the Cambridge Structural Database: deposition number 1429949. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L.S. (stuart_schreiber@harvard.edu).

METHODS

In vitro *P. falciparum* blood-stage culture and assay. Strains of *P. falciparum* (Dd2, 3D7, D6, K1, NF54, V1/3, HB3, 7G8, FCB and TM90C2B) were obtained from the Malaria Research and Reference Reagent Resource Center (MR4). *PfscDHODH*, the transgenic *P. falciparum* line expressing *S. cerevisiae* DHODH¹⁹, was a gift from A. B. Vaidya. *P. falciparum* isolates were maintained with O-positive human blood in an atmosphere of 93% N₂, 4% CO₂, 3% O₂ at 37°C in complete culturing medium (10.4 g l⁻¹ RPMI 1640, 5.94 g l⁻¹ HEPES, 5 g l⁻¹ albumax II, 50 mg l⁻¹ hypoxanthine, 2.1 g l⁻¹ sodium bicarbonate, 10% human serum and 43 mg l⁻¹ gentamicin). Parasites were cultured in medium until parasitaemia reached 3–8%. Parasitaemia was determined by checking at least 500 red blood cells from a Giemsa-stained blood smear. For the compound screening, a parasite dilution at 2.0% parasitaemia and 2.0% haematocrit was created with medium. 25 µl of medium was dispensed into 384-well black, clear-bottom plates and 100 nl of each compound in DMSO was transferred into assay plates along with the control compound (mefloquine). Next, 25 µl of the parasite suspension in medium was dispensed into the assay plates giving a final parasitaemia of 1% and a final haematocrit of 1%. The assay plates were incubated for 72 h at 37°C. 10 µl of detection reagent consisting of 10 × SYBR Green I (Invitrogen; supplied in 10,000 × concentration) in lysis buffer (20 mM Tris-HCl, 5 mM EDTA, 0.16% (w/v) Saponin, 1.6% (v/v) Triton X-100) was dispensed into the assay plates. For optimal staining, the assay plates were left at room temperature for 24 h in the dark. The assay plates were read with 505 dichroic mirrors with 485 nm excitation and 530 nm emission settings in an Envision (PerkinElmer).

Chemoinformatics clustering. High-throughput screening hits were hierarchically clustered by structural similarity using average linkage on pairwise Jaccard distances⁴³ between ECFP4 fingerprints⁴⁴. Pipeline Pilot⁴⁵ was used for fingerprint and distance calculation; clustering and heat-map generation was done in R (ref. 46).

In vitro *P. berghei* liver-stage assay. HepG2 cells (ATCC) were maintained in DMEM, 10% (v/v) FBS (Sigma), and 1% (v/v) antibiotic-antimycotic in a standard tissue culture incubator (37°C, 5% CO₂). *P. berghei* (ANKA GFP-luc) infected *A. stephensi* mosquitoes were obtained from the New York University Langone Medical Center Insectary. For assays, ~17,500 HepG2 cells per well were added to a 384-well microtitre plate in duplicate. After 18–24 h at 37°C the media was exchanged and compounds were added. After 1 h, parasites obtained from freshly dissected mosquitoes were added to the plates (4,000 parasites per well), the plates were spun for 10 min at 1,000 r.p.m. and then incubated at 37°C. The final assay volume was 30 µl. After a 48-h incubation at 37°C, Bright-Glo (Promega) was added to the parasite plate to measure relative luminescence. The relative signal intensity of each plate was evaluated with an EnVision (PerkinElmer) system.

In vitro *P. falciparum* liver-stage assay. Micropatterned co-culture (MPCC) is an *in vitro* co-culture system of primary human hepatocytes organized into colonies and surrounded by supportive stromal cells. Hepatocytes in this format maintain a functional phenotype for up to 4–6 weeks without proliferation, as assessed by major liver-specific functions and gene expression^{47–49}. In brief, 96-well plates were coated homogeneously with rat-tail type I collagen (50 µg ml⁻¹) and subjected to soft-lithographic techniques to pattern the collagen into 500-µm-island microdomains that mediate selective hepatocyte adhesion. To create MPCCs, cryopreserved primary human hepatocytes (BioreclamationIVT) were pelleted by centrifugation at 100g for 6 min at 4°C, assessed for viability using Trypan blue exclusion (typically 70–90%), and seeded on micropatterned collagen plates (each well contained ~10,000 hepatocytes organized into colonies of 500 µm) in serum-free DMEM with 1% penicillin–streptomycin. The cells were washed with serum-free DMEM with 1% penicillin–streptomycin 2–3 h later and replaced with human hepatocyte culture medium⁴⁸. 3T3-J2 mouse embryonic fibroblasts were seeded (7,000 cells per well) 24 h after hepatocyte seeding. 3T3-J2 fibroblasts were courtesy of H. Green⁵⁰.

MPCCs were infected with 75,000 sporozoites (NF54) (Johns Hopkins University) 1 day after hepatocytes were seeded^{48,49}. After incubation at 37°C and 5% CO₂ for 3 h, wells were washed once with PBS, and the respective compounds were added. Cultures were dosed daily. Samples were fixed on day 3.5 after infection. For immunofluorescence staining, MPCCs were fixed with –20°C methanol for 10 min at 4°C, washed twice with PBS, blocked with 2% BSA in PBS, and incubated with mouse anti-*P. falciparum* Hsp70 antibodies (clone 4C9, 2 µg ml⁻¹) for 1 h at room temperature. Samples were washed with PBS then incubated with Alexa 488-conjugated secondary goat anti-mouse for 1 h at room temperature. Samples were washed with PBS, counterstained with the DNA dye Hoechst 33258 (Invitrogen; 1:1,000), and mounted on glass slides with fluoromount G (Southern Biotech). Images were captured on a Nikon Eclipse Ti fluorescence microscope. Diameters of developing liver stage parasites were measured and used to calculate the corresponding area.

In vitro *P. cynomolgi* liver-stage assay. All rhesus macaques (*Macaca mulatta*) used in this study were bred in captivity for research purposes, and were housed

at the Biomedical Primate Research Centre (BPRC; AAALAC-certified institute) facilities under compliance with the Dutch law on animal experiments, European directive 86/609/EEC and with the ‘Standard for Humane Care and Use of Laboratory Animals by Foreign Institutions’ identification number A5539-01, provided by the Department of Health and Human Services of the US National Institutes of Health. The local independent ethical committee first approved all protocols. Non-randomized rhesus macaques (male or female; 5–14 years old; one animal per month) were infected with 1 × 10⁶ *P. cynomolgi* (M strain) blood-stage parasites, and bled at peak parasitaemia. Approximately 300 female *A. stephensi* mosquitoes (Sind-Kasur strain, Nijmegen University Medical Centre St Radboud) were fed with this blood as described previously⁵¹.

Rhesus monkey hepatocytes were isolated from liver lobes as described by previously⁵². Sporozoite infections were performed within 3 days of hepatocyte isolation. Sporozoite inoculation of primary rhesus monkey hepatocytes was performed as described previously^{53,54}. On day 6, intracellular *P. cynomolgi* malaria parasites were fixed, stained with purified rabbit antiserum reactive against *P. cynomolgi* Hsp70.1 (ref. 53), and visualized with FITC-labelled goat anti-rabbit IgG antibodies. Quantification of small ‘hypnozoite’ exoerythrocytic forms (1 nucleus, a small round shape, a maximal diameter of 7 µm) or large ‘developing parasite’ exoerythrocytic forms (more than 1 nucleus, larger than 7 µm and round or irregular shape) was determined for each well using a high-content imaging system (Operetta, PerkinElmer).

In vitro transmission-blocking assay (gametocyte IV–V). *P. falciparum* 3D7 stage IV–V gametocytes were isolated by discontinuous Percoll gradient centrifugation of parasite cultures treated with 50 mM N-acetyl-D-glucosamine for 3 days to kill asexual parasites. Gametocytes (1.0 × 10⁵) were seeded in 96-well plates and incubated with compounds for 72 h. *In vitro* anti-gametocyte activity was measured using CellTiter-Glo (Promega).

In vitro transmission-blocking imaging assay (early, I–III; and late, IV–V, gametocyte). A detailed description of the method is published elsewhere⁵⁵. In brief, NF54^{pf16-LUC-GFP} highly synchronous gametocytes were induced from a single intra-erythrocytic asexual replication cycle. On day 0 of gametocyte development, spontaneously generated gametocytes were removed by VarioMACS magnetic column (MAC) technology. Early stage I gametocytes were collected on day 2 of development and late-stage gametocytes (stage IV) on day 8 using MAC columns. Percentage parasitaemia and haematocrit was adjusted to 10 and 0.1, respectively. 45 µl of parasite sample were added to PerkinElmer Cell carrier poly-D-lysine imaging plates containing 5 µl of test compound at 16 doses, including control wells containing 4% DMSO and 50 µM puromycin (0.4% and 5 µM final concentrations, respectively), the plates sealed with a membrane (Breatheasy or 4ti-05 15/ST) and incubated for 72 h in standard incubation conditions of 5% CO₂, 5% O₂, 90% N₂ and 60% humidity at 37°C. After incubation, 5 µl of 0.07 µg ml⁻¹ MitoTracker Red CM-H2XRos (MTR) (Invitrogen) in PBS was added to each well, and plates were resealed with membranes and incubated overnight under standard conditions. The following day, the plates were brought to room temperature for at least one hour before being measured on the Opera QEHs Instrument. Image analysis was performed using an Acapella (PerkinElmer)-based algorithm that identifies gametocytes of the expected morphological shape with respect to degree of elongation and specifically those parasites that are determined as viable by the MitoTracker Red CM-H2XRos fluorescence size and intensity. IC₅₀ values were determined using GraphPad Prism 4, using a 4-parameter log dose, nonlinear regression analysis, with sigmoidal dose–response (variable slope) curve fit.

***P. falciparum* standard membrane feeding assay.** *P. falciparum* transmission-blocking activity of BRD7929 was assessed in a standard membrane feeding assay as previously described⁵⁶. In brief, *P. falciparum*^{NF54 hsp70-GFP-luc} reporter parasites were cultured up to stage V gametocytes (day 14). Test compounds were serially diluted in DMSO and subsequently in RPMI medium to reach a final DMSO concentration of 0.1%. Diluted compound was either pre-incubated with stage V gametocytes for 24 h (indirect mode) or directly added to the blood meal (direct mode). Gametocytes were adjusted to 50% haematocrit, 50% human serum and fed to *A. stephensi* mosquitoes. All compound dilutions were tested in duplicate in independent feeders. After 8 days, mosquitoes were collected and the relative decrease in oocyst density in the midgut was determined by measurement of luminescence signals in 24 individual mosquitoes from each cage. For each vehicle (control) cage, an additional 10 mosquitoes were dissected and examined by microscopy to determine the baseline oocyst intensity.

In vitro resistance selections. *In vitro* resistance selections were performed as previously described¹⁵. In brief, approximately 1 × 10⁹ *P. falciparum* Dd2 parasites were treated with 60 nM (EC_{99.9}) or 150 nM (10 × EC₅₀) of BRD1095 in each of four independent flasks for 3–4 days. After the compounds were removed, the cultures were maintained in compound-free complete RPMI growth medium with regular media exchange until healthy parasites reappeared. Once parasitaemia reached

2–4%, compound pressure was repeated and these steps were executed for about 2 months until the initial EC₅₀ shift was observed. Three out of four independent selections pressured at 60 nM developed a phenotypic EC₅₀ shift. None of the selections pressured at 150 nM resulted in resistant parasites. After an initial shift in the dose–response phenotype was observed, selection at an increased concentration was repeated in the same manner until at least a threefold shift in EC₅₀ was observed. Selected parasites were then cloned by limiting dilution.

BRD73842-resistant selections were conducted in a similar manner except that parasites were initially treated at 0.5 μ M (10 \times EC₅₀) for 4 days or 150 nM (EC_{99.9}) for 2 days in each of two independent flasks. The Y1356N mutant was derived from a flask pressured at 0.5 μ M and the L1418F mutant was developed from one of the flasks exposed to the 150 nM.

Whole-genome sequencing and target identification. DNA libraries were prepared for sequencing using the Illumina Nextera XT kit (Illumina), and quality-checked before sequencing on a TapeStation. Libraries were clustered and run as 100-bp paired-end reads on an Illumina HiSeq 2000 in RapidRun mode, according to the manufacturer's instructions. Samples were analysed by aligning to the *P. falciparum* 3D7 reference genome (PlasmoDB v. 11.1). To identify SNVs and CNVs, a sequencing pipeline developed for *P. falciparum* (Plasmodium Type Uncovering Software, Platypus) was used as previously described, with the exception of an increase in the base quality filter from 196.5 to 1,000 (ref. 57).

***P. falciparum* DHODH biochemical assay.** Substrate-dependent inhibition of recombinant *P. falciparum* DHODH protein was assessed in an *in vitro* assay in 384-well clear plates (Corning 3640) as described previously⁵⁸. A 20-point dilution series of inhibitor concentrations were assayed against 2–10 nM protein with 500 μ M L-dihydroorotate substrate (excess), 18 μ M dodecylubiquinone electron acceptor ($\sim K_m$), and 100 μ M 2,6-dichloroindophenol indicator dye in assay buffer (100 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol, 0.5% Triton X-100). Assays were incubated at 25 °C for 20 min and then assessed via OD₆₀₀. Data were normalized to 1% DMSO and excess inhibitor (25 μ M DSM265; ref. 7).

Human DHODH biochemical assay. Substrate-dependent inhibition of recombinant human DHODH protein was assessed in an *in vitro* assay in 384-well clear plates (Corning 3640) as described previously⁵⁹. A 20-point dilution series of inhibitor concentrations was assayed against 13 nM protein with 1 mM L-dihydroorotate substrate (excess), 100 μ M dodecylubiquinone electron acceptor, and 60 μ M 2,6-dichloroindophenol indicator dye in assay buffer (50 mM Tris HCl pH 8.0, 150 mM KCl, 0.1% Triton X-100). Assays were incubated at 25 °C for 20 min and then assessed via OD₆₀₀. Data were normalized to 1% DMSO and no enzyme.

***P. vivax* PI4K biochemical assay.** The synthetic gene for full-length *P. vivax* PI4K (PVX_098050) was synthesized from GeneArt (ThermoScientific), and was expressed and purified as previously described²⁰. Aliquots of *P. vivax* PI4K β were flash-frozen in liquid nitrogen and stored at –80 °C. Full-length human PI4KB (uniprot gene Q9UBF8-2) was expressed and purified as previously described⁶⁰. 100 nM extruded lipid vesicles were made to mimic Golgi organelle vesicles (20% phosphatidylinositol, 10% phosphatidylserine, 45% phosphatidylcholine and 25% phosphatidylethanolamine) in lipid buffer (20 mM HEPES pH 7.5 (room temperature), 100 mM KCl, 0.5 mM EDTA). Lipid kinase assays were carried out using the Transcreener ADP² FI Assay (BellBrook Labs) following the published protocol as previously described⁶¹. 4- μ l reactions ran at 21 °C for 30 min in a buffer containing 30 mM HEPES pH 7.5, 100 mM NaCl, 50 mM KCl, 5 mM MgCl₂, 0.25 mM EDTA, 0.4% (v/v) Triton X-100, 1 mM TCEP, 0.5 mg ml^{–1} Golgi-mimic vesicles and 10 μ M ATP. *P. vivax* PI4K β was used at 7.5 nM and human PI4KB was used at 200 nM. Fluorescence intensity was measured using a Spectramax M5 plate reader with excitation at 590 nm and emission at 620 nm (20-nm bandwidth). IC₅₀ values were calculated from triplicate inhibitor curves using GraphPad Prism software.

PheRS homology modelling. The model was built using the SWISS-MODEL online resource^{62–64} and Prime⁶⁵ (Schrödinger Release 2015-2: Prime, version 4.0, Schrödinger), with human PheRS (PDB accession 3L4G) as a template for *P. falciparum* PheRS (PlasmoDB Gene ID: PF3D7_0109800). The template was chosen based on highest sequence identity and similarity identified via PSI-BLAST. Target-template alignment was made using ProMod-II and validated with Prime STA. Coordinates from residues that were conserved between the target and the template were copied from the template to the model, and remaining sites were remodelled using segments from known structures. The side chains were then rebuilt, and the model was finally refined using a force field.

***P. falciparum* cytoplasmic PheRS biochemical assay.** Protein sequences of both α - (PF3D7_0109800) and β - (PF3D7_1104000) subunits of cytoplasmic *P. falciparum* PheRS were obtained from PlasmoDB (<http://plasmodb.org/plasmo/>). Full length α - and β -subunit gene sequences optimized for expression in *E. coli* were cloned into pETM11 (Kanamycin resistance) and pETM20 (ampicillin resistance) expression vectors using NcoI and KpnI sites and co-transformed into *E. coli* B834 cells. Protein expression was induced by addition of 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and cells were grown until an OD₆₀₀ of

0.6–0.8 was reached at 37 °C. They were then allowed to grow at 18 °C for 20 h after induction. Cells were separated by centrifugation at 5,000g for 20 min and the bacterial pellets were suspended in a buffer consisting of 50 mM Tris–HCl (pH 7.5), 200 mM NaCl, 4 mM β -mercaptoethanol, 15% (v/v) glycerol, 0.1 mg ml^{–1} lysozyme and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were lysed by sonication and cleared by centrifugation at 20,000g for 1 h. The supernatant was applied on to prepacked NINTA column (GE Healthcare), and bound proteins were eluted by gradient-mixing with elution buffer (50 mM Tris–HCl (pH 7.5), 80 mM NaCl, 4 mM β -mercaptoethanol, 15% (v/v) glycerol, 1 M imidazole). Pure fractions were pooled and loaded on to heparin column for further purification. Again, bound proteins were eluted using gradient of heparin elution buffer 50 mM Tris–HCl (pH 7.5), 1 M NaCl, 4 mM β -mercaptoethanol, 15% (v/v) glycerol). Pure fractions were again pooled and dialysed overnight into a buffer containing 50 mM Tris–HCl (pH 7.5), 200 mM NaCl, 4 mM β -mercaptoethanol, 1 mM DTT and 0.5 mM EDTA. TEV protease (1:50 ratio of protease:protein) was added to the protein sample and incubated at 20 °C for 24 h to remove the polyhistidine tag. Protein was further purified via gel-filtration chromatography on a GE HiLoad 60/600 Superdex column in 50 mM Tris–HCl (pH 7.5), 200 mM NaCl, 4 mM β -mercaptoethanol, 1 mM MgCl₂. The eluted protein (a heterodimer of *P. falciparum* cPheRS) were collected, assessed for purity via SDS–PAGE and stored at –80 °C.

Nuclear encoded tRNA^{Phe} from *P. falciparum* was synthesized using an *in vitro* transcription method as described earlier^{22,66}. Aminoacylation and enzyme inhibition assays for *P. falciparum* cytosolic PheRS were performed as described earlier^{22,67}. Enzymatic assays were performed in buffer containing 30 mM HEPES (pH 7.5), 150 mM NaCl, 30 mM KCl, 50 mM MgCl₂, 1 mM DTT, 100 μ M ATP, 100 μ M L-phenylalanine, 15 μ M *P. falciparum* tRNA^{Phe}, 2 U ml^{–1} *E. coli* inorganic pyrophosphatase (NEB) and 500 nM recombinant *P. falciparum* PheRS at 3 °C. Reactions at different time points were stopped by the addition of 40 mM EDTA and subsequent transfer to ice. Recombinant maltose binding protein was used as negative control. The cPheRS inhibition assays were performed using inhibitor concentrations of 0.01 nM, 0.1 nM, 1 nM, 10 nM, 100 nM, 1 μ M, 5 μ M and 10 μ M for strong binders and 1 nM, 10 nM, 100 nM, 1 μ M, 10 μ M, 100 μ M and 500 μ M for weaker binders in the assay buffer. Enzymatic and inhibition experiments were performed twice in triplicate.

Mammalian cell cytotoxicity assays. Mammalian cells (HepG2, A549, and HEK293) were obtained from the ATCC and cultured routinely in DMEM with 10% FBS and 1% (v/v) antibiotic–antimycotic. For cytotoxicity assays, 1 \times 10⁶ cells were seeded into 384-well plates 1 day before compound treatment. Cells were treated with ascending doses of compound for 72 h, and viability was measured using Cell-Titer Glo (Promega). All cell lines were tested for *Mycoplasma* contamination using Universal mycoplasma Detection Kit (ATCC).

***In vitro* ADME/PK and safety assays.** *In vitro* characterization assays (protein binding, microsomal stability, hepatocyte stability, cytochrome P450 (CYP) inhibition, and aqueous solubility) were performed according to industry-standard techniques. Ion channel inhibition studies were performed using the Q-Patch system using standard techniques.

Animal welfare. All animal experiments were conducted in compliance with institutional policies and appropriate regulations and were approved by the institutional animal care and use committees for each of the study sites (the Broad Institute, 0016-09-14; Harvard School of Public Health, 03228; Eisai, 13-05, 13-07, 14-C-0027). No method of randomization or blinding was used in this study.

***In vivo P. berghei* blood-stage assay.** CD-1 mice (n = 4 per experimental group; female; 6–7-week-old; 20–24 g, Charles River) were intravenously inoculated with approximately 1 \times 10⁵ *P. berghei* (ANKA GFP-luc) blood-stage parasites 24 h before treatment and compounds were administered orally (at 0 h). Parasitaemia was monitored by the *in vivo* imaging system (IVIS SpectrumCT, PerkinElmer) to acquire the bioluminescence signal (150 mg kg^{–1} of luciferin was intraperitoneally injected approximately 10 min before imaging). In addition, blood smear samples were obtained from each mouse periodically, stained with Giemsa, and viewed under a microscope for visual detection of blood parasitaemia. Animals with parasitaemia exceeding 25% were humanely euthanized.

***In vivo P. berghei* causal prophylaxis assay.** CD-1 mice (n = 4 per experimental group; female; 6–7-week-old; 20–24 g, Charles River) were inoculated intravenously with approximately 1 \times 10⁵ *P. berghei* (ANKA GFP-luc) sporozoites freshly dissected from *A. stephensi* mosquitoes. Immediately after infection, the mice were treated with single oral doses of BRD7929; infection was monitored as described for the *P. berghei* erythrocytic-stage assay. For time-course experiments, the time of compound treatment (single oral dose of 10 mg kg^{–1}) was varied from 5 days before infection to 2 days after infection.

***In vivo P. berghei* transmission-stage assay.** CD-1 (n = 3 per experimental group; female; 6–7-week-old; 21–24 g, Charles River) mice were infected with *P. berghei* (ANKA GFP-luc) for 96 h before treatment with vehicle or BRD7929 (day 0).

On day 2, female *A. stephensi* mosquitoes were allowed to feed on the mice for 20 min. After 1 week (day 9), the midguts of the mosquitoes were dissected out and oocysts were enumerated microscopically (12.5× magnification).

In vivo *P. falciparum* blood-stage assay. *In vivo* adapted *P. falciparum* (3D7^{HLH/BRD}) were selected as described previously⁶⁸. In brief, NSG mice ($n = 2$ per experimental group; female; 4–5-week-old; 19–21 g; The Jackson Laboratory) were intraperitoneally injected with 1 ml of human erythrocytes (O-positive, 50% haematocrit, 50% RPMI 1640 with 5% albumax) daily to generate mice with humanized circulating erythrocytes (huRBC NSG). Approximately 2×10^7 blood-stage *P. falciparum* 3D7^{HLH/BRD} (ref. 69) were intravenously infected to huRBC NSG mice and >1% parasitaemia was achieved 5 weeks after infection. After three *in vivo* passages, the parasites were frozen and used experimentally.

Approximately 48 h after infection with 1×10^7 blood-stage of *P. falciparum* 3D7^{HLH/BRD}, the mean parasitaemia was approximately 0.4%. huRBC NSG mice were orally treated with a single dose of compound and parasitaemia was monitored for 30 days by IVIS to acquire the bioluminescence signal (150 mg kg⁻¹ of luciferin was intraperitoneally injected approximately 10 min before imaging).

In vivo *P. falciparum* transmission-stage assay. huRBC NSG mice ($n = 2$ per experimental group; female; 4–5-week-old; 18–20 g; Jackson Laboratory) were infected with blood-stage *P. falciparum* 3D7^{HLH/BRD} for 2 weeks to allow the development of mature gametocytes. Subsequently, the mice were treated with a single oral dose of BRD7929. Blood samples were collected for 11 days. For molecular detection of parasite stages, 40 µl of blood was obtained from control and treated mice. In brief, total RNA was isolated from blood samples using RNeasy Plus Kit with genomic DNA eliminator columns (Qiagen). First-strand cDNA synthesis was performed from extracted RNA using SuperScript III First-Strand Synthesis System (Life Technologies). Parasite stages were quantified using a stage-specific qRT-PCR assay as described previously^{33,69}. Primers were designed to measure transcript levels of PF3D7_0501300 (ring stage parasites), PF3D7_1477700 (immature gametocytes) and PF3D7_1031000 (mature gametocytes). Primers for PF3D7_1120200 (*P. falciparum* UCE) transcript were used as a constitutively expressed parasite marker. The assay was performed using cDNA in a total reaction volume of 20 µl, containing primers for each gene at a final concentration of 250 nM. Amplification was performed on a Vii7 qRT-PCR machine (Life Technologies) using SYBR Green Master Mix (Applied Biosystems) with the following reaction conditions: 1 cycle \times 10 min at 95°C and 40 cycles \times 1 s at 95°C and 20 s at 60°C. Each cDNA sample was run in triplicate and the mean C_t value was used for the analysis. C_t values obtained above the cut-off (negative control) for each marker were considered negative for the presence of specific transcripts. Blood samples from each mouse before parasite inoculation were also tested for 'background noise' using the same primer sets. No amplification was detected from any samples.

In vivo *P. falciparum* liver-stage assay. FRG knockout on C57BL/6 (human repopulated, >70%) mice (huHep FRG knockout; $n = 2$ per experimental group; female; 5.5–6-month-old; 19–21 g; Yecuris) were inoculated intravenously with approximately 1×10^5 *P. falciparum* (NF54HT-GFP-luc) sporozoites and BRD7929 was administered as a single 10 mg kg⁻¹ oral dose one day after inoculation³¹. Infection was monitored daily by IVIS. Daily engraftment of human erythrocytes (0.4 ml, O-positive, 50% haematocrit, 50% RPMI 1640 with 5% albumax) was initiated 5 days after inoculation. For qPCR analysis, blood samples (40 µl) were collected 7 days after inoculation. For molecular detection of the blood-stage parasite, 40 µl of blood was obtained from control and treated mice. In brief, total RNA was isolated from blood samples using RNeasy Plus Kit with genomic DNA eliminator columns (Qiagen). First-strand cDNA synthesis was performed from extracted RNA using SuperScript III First-Strand Synthesis System (Life Technologies). The presence of the blood-stage parasites was quantified using a highly stage-specific qRT-PCR assay as described previously^{33,70}. Primers were designed to measure transcript levels of PF3D7_1120200 (*P. falciparum* UCE). The assay was performed using cDNA in a 20 µl total reaction volume containing primers for each gene at a final concentration of 250 nM. Amplification was performed on a Vii7 qRT-PCR machine (Life Technologies) using SYBR Green Master Mix (Applied Biosystems) and the reaction conditions are as follows: 1 cycle \times 10 min at 95°C and 40 cycles \times 1 s at 95°C and 20 s at 60°C. Each cDNA sample was run in triplicate and the mean C_t value was used for the analysis. C_t values obtained above the cut-off (negative control) for each marker were considered negative for presence of specific transcripts. Blood samples from each mouse were also tested for background noise using the same primer sets before parasite inoculation. No amplification was detected from any samples.

Resistance propensity determination assay. *In vitro* cultures of *P. falciparum* Dd2, with the initial inocula ranging from 10^5 to 10^9 parasites, were maintained in complete medium supplemented with 20 nM of BRD7929 (EC₉₀ against Dd2). Media was replaced with fresh compound added daily and cultures monitored

for 60 days to identify propensity for recrudescence parasitaemia as described³⁴. Atovaquone was used as a control (EC₉₀ = 2 nM).

Solubility assay. Solubility was determined in PBS pH 7.4 with 1% DMSO. Each compound was prepared in triplicate at 100 µM in both 100% DMSO and PBS with 1% DMSO. Compounds were allowed to equilibrate at room temperature with a 750 r.p.m. vortex shake for 18 h. After equilibration, samples were analysed by UPLC-MS (Waters) with compounds detected by single-ion reaction detection on a single quadrupole mass spectrometer. The DMSO samples were used to create a two-point calibration curve to which the response in PBS was fit.

Plasma protein binding assay. Plasma protein binding was determined by equilibrium dialysis using the Rapid Equilibrium Dialysis (RED) device (Pierce Biotechnology) for both human and mouse plasma. Each compound was prepared in duplicate at 5 µM in plasma (0.95% acetonitrile, 0.05% DMSO) and added to one side of the membrane (200 µl) with PBS pH 7.4 added to the other side (350 µl). Compounds were incubated at 37°C for 5 h with 350 r.p.m. orbital shaking. After incubation, samples were analysed by UPLC-MS (Waters) with compounds detected by SIR detection on a single quadrupole mass spectrometer.

hERG channel inhibition assay. The required potency to inhibit the hERG channel in expressed cell lines were evaluated using an automated patch-clamp system (QPatch-HTX).

Mouse pharmacokinetics assay. Pharmacokinetics of BRD3444 and BRD1095 were performed by Shanghai ChemPartner Co. Ltd., following single intravenous and oral administrations to female CD-1 mice. BRD3444 and BRD1095 were formulated in 70% PEG400 and 30% aqueous glucose (5% in H₂O) for intravenous and oral dosing. Test compounds were dosed as a bolus solution intravenously at 0.6 mg kg⁻¹ (dosing solution; 70% PEG400 and 30% aqueous glucose, 5% in H₂O) or dosed orally by gavage as a solution at 1 mg kg⁻¹ (dosing solution; 70% PEG400 and 30% aqueous glucose, 5% in H₂O) to female CD-1 mice ($n = 9$ per dose route). Pharmacokinetic parameters of BRD7929 and BRD3316 were determined in CD-1 mice. BRD7929 and BRD3316 were formulated in 10% ethanol, 4% Tween, 86% saline for both intravenous and oral dosing. Pharmacokinetic parameters were estimated by non-compartmental model using WinNonlin 6.2. Pharmacokinetic parameters for BRD7929 and BRD3316 were estimated by a non-compartmental model using proprietary Eisai software. Pharmacokinetic parameters of BRD7539 and BRD9185 were determined in CD-1 mice. Compounds were formulated in 70% PEG300 and 30% (5% glucose in H₂O) at 0.5 mg ml⁻¹ for oral dosing, and 5% DMSO, 10% cremophor, and 85% H₂O at 0.25 mg ml⁻¹ for intravenous formulation. Pharmacokinetic parameters were estimated by non-compartmental model using WinNonlin 6.2. Pharmacokinetics of BRD7539 and BRD9185 were performed by WuXi AppTec. The protocol was approved by Eisai IACUC, 13-07, 13, 05, and 14-c-0027.

Metabolic stability assay. Compounds were evaluated *in vitro* to determine their metabolic stability in incubations containing liver microsomes or hepatocytes of mouse and human. In the presence of NADPH, liver microsomes (0.2 mg ml⁻¹) from mouse (CD-1) and human were incubated with compounds (0.5 and 5 µM) for 0, 10 and 90 min. The depletion of compounds in the incubation mixtures, determined using liquid chromatography tandem mass spectrometry LC-MS/MS, was used to estimate K_m and V_{max} values and determine half-lives for both mouse and human microsomes.

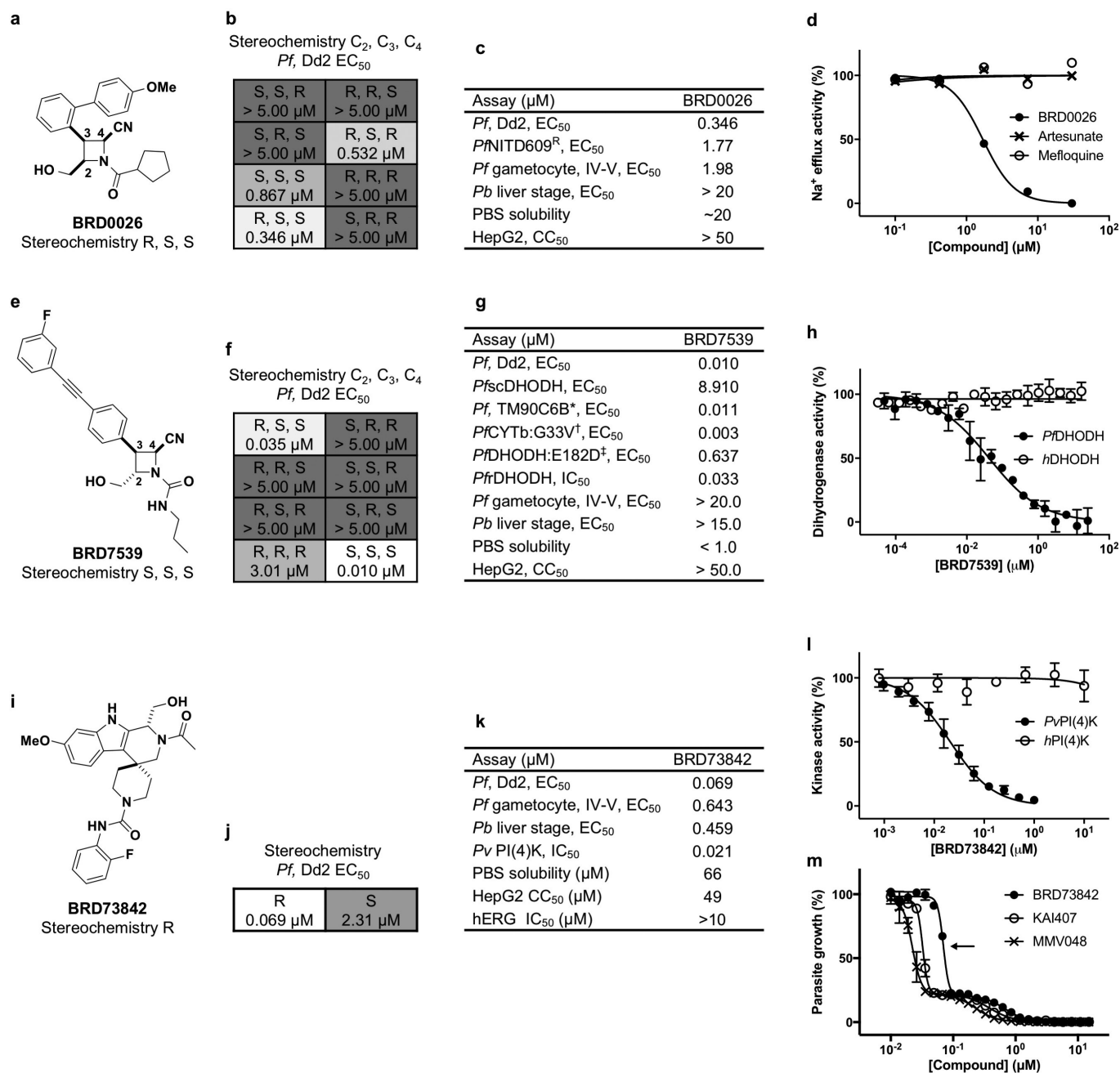
CYP inhibition assay. Compounds were evaluated *in vitro* for the potential inhibition of human cytochrome P450 (CYP) isoforms using human liver microsomes. Two concentrations (1 and 10 µM) of compound were incubated with pooled liver microsomes (0.2 mg ml⁻¹) and a cocktail mixture of probe substrates for selective CYP isoform. The selective activities tested were CYP1A2-mediated phenacetin O-demethylation, CYP2C8-mediated rosiglitazone para-hydroxylation, CYP2C9-mediated tolbutamide 4'-hydroxylation, CYP2C19-mediated (S)-mephenytoin 4'-hydroxylation, CYP2D6-mediated (±)-bupropion 1'-hydroxylation and, CYP3A4/5-mediated midazolam 1'-hydroxylation. The positive controls tested were α-naphthoflavone for CYP1A2, montelukast for CYP2C8, sulfaphenazole for CYP2C9, tranylcypromine for CYP2C19, quinidine for CYP2D6, and ketoconazole for CYP3A4/5. The samples were analysed by LC-MS/MS. IC₅₀ values were estimated using nonlinear regression.

Time-dependent inactivation assay. The time-dependent inactivation potential of compounds were assessed in human liver microsomes for CYP2C9, CYP2D6, and CYP3A4/5 by determining K_I and k_{inact} values when appropriate. Two concentrations (6 and 30 µM) of compound were incubated in primary reaction mixtures containing phosphate buffer and 0.2 mg ml⁻¹ human liver microsomes for 0, 5, and 30 min in a 37°C water bath. The reactions were initiated by the addition of NADPH. Phosphate buffer was substituted for NADPH solution for control. At the respective times, 25 µl of primary incubation was diluted tenfold into pre-incubated secondary incubation mixture containing each CYP-selective probe substrate in order to assess residual activity. The second incubation time was 10 min. The probe substrates used for CYP1A, 2C9, CYP2C19, CYP2D6,

and CYP3A4 were phenacetin (50 μ M), tolbutamide (500 μ M), (S)-mephénytoin (20 μ M), bufuralol (50 μ M), and midazolam (30 μ M), respectively. The CYP time-dependent inhibitors used were furafyllin, tienilic acid, ticlopidine, paroxetine and troleandomycin for CYP2C8, CYP2C9, CYP2C19, CYP2D6 and CYP3A, respectively, at two concentrations. The samples were analysed by LC–MS/MS.

Chemical synthesis and analytical data. See Supplementary Methods.

43. Jaccard, P. Lois de distribution florale dans la zone alpine. *Bull. Soc. Vaud. Sci. Nat.* **38**, 69–130 (1902).
44. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
45. Pipeline Pilot v. 8.5.0.200; <http://cscenter.pbsci.ucsc.edu:9944/> (Accelrys Software Inc., 2011).
46. R Core Development Team. *R: A Language and Environment for Statistical Computing* v. 3.0; <http://www.R-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2013).
47. Khetani, S. R. & Bhatia, S. N. Microscale culture of human liver cells for drug development. *Nat. Biotechnol.* **26**, 120–126 (2008).
48. March, S. *et al.* A microscale human liver platform that supports the hepatic stages of *Plasmodium falciparum* and *vivax*. *Cell Host Microbe* **14**, 104–115 (2013).
49. March, S. *et al.* Micropatterned coculture of primary human hepatocytes and supportive cells for the study of hepatotropic pathogens. *Nat. Protocols* **10**, 2027–2053 (2015).
50. Rheinwald, J. G. & Green, H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* **6**, 331–343 (1975).
51. Ponnudurai, T. *et al.* Infectivity of cultured *Plasmodium falciparum* gametocytes to mosquitoes. *Parasitology* **98**, 165–173 (1989).
52. Guguen-Guillouzo, C. *et al.* High yield preparation of isolated human adult hepatocytes by enzymatic perfusion of the liver. *Cell Biol. Int. Rep.* **6**, 625–628 (1982).
53. Dembele, L. *et al.* Towards an *in vitro* model of *Plasmodium* hypnozoites suitable for drug discovery. *PLoS One* **6**, e18162 (2011).
54. Mazier, D. *et al.* Complete development of hepatic stages of *Plasmodium falciparum* *in vitro*. *Science* **227**, 440–442 (1985).
55. Duffy, S. & Avery, V. M. Identification of inhibitors of *Plasmodium falciparum* gametocyte development. *Malar. J.* **12**, 408 (2013).
56. Stone, W. J. R. *et al.* A scalable assessment of *Plasmodium falciparum* transmission in the standard membrane-feeding assay, using transgenic parasites expressing green fluorescent protein-luciferase. *J. Infect. Dis.* **210**, 1456–1463 (2014).
57. Manary, M. J. *et al.* Identification of pathogen genomic variants through an integrated pipeline. *BMC Bioinformatics* **15**, 1–14 (2014).
58. Ross, L. S. *et al.* *In vitro* resistance selections for *Plasmodium falciparum* dihydroorotate dehydrogenase inhibitors give mutants with multiple point mutations in the drug-binding site and altered growth. *J. Biol. Chem.* **289**, 17980–17995 (2014).
59. Hansen, M. *et al.* Inhibitor binding in a class 2 dihydroorotate dehydrogenase causes variations in the membrane-associated N-terminal domain. *Protein Sci.* **13**, 1031–1042 (2004).
60. Burke, J. E. *et al.* Structures of PI4KIII β complexes show simultaneous recruitment of Rab11 and its effectors. *Science* **344**, 1035–1038 (2014).
61. Burke, J. E. & Williams, R. L. Dynamic steps in receptor tyrosine kinase mediated activation of class IA phosphoinositide 3-kinases (PI3K) captured by H/D exchange (HDX-MS). *Adv. Biol. Regul.* **53**, 97–110 (2013).
62. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
63. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).
64. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
65. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367 (2004).
66. Sherlin, L. D. *et al.* Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA* **7**, 1671–1678 (2001).
67. Cestari, I. & Stuart, K. A spectrophotometric assay for quantitative measurement of aminoacyl-tRNA synthetase activity. *J. Biomol. Screen.* **18**, 490–497 (2013).
68. Angulo-Barturen, I. *et al.* A murine model of *falciparum*-malaria by *in vivo* selection of competent strains in non-myelodepleted mice engrafted with human erythrocytes. *PLoS One* **3**, e2252 (2008).
69. Ekland, E. H., Schneider, J. & Fidock, D. A. Identifying apicoplast-targeting antimalarials using high-throughput compatible approaches. *FASEB J.* **25**, 3583–3593 (2011).
70. Joice, R. *et al.* Inferring developmental stage composition from gene expression in human malaria. *PLOS Comput. Biol.* **9**, e1003392 (2013).
71. Ruecker, A. *et al.* A male and female gametocyte functional viability assay to identify biologically relevant malaria transmission-blocking drugs. *Antimicrob. Agents Chemother.* **58**, 7292–7302 (2014).
72. Zeeman, A.-M. *et al.* KAI407, a potent non-8-aminoquinoline compound that kills *Plasmodium cynomolgi* early dormant liver stage parasites *in vitro*. *Antimicrob. Agents Chemother.* **58**, 1586–1595 (2014).
73. Lukens, A. K. *et al.* Harnessing evolutionary fitness in *Plasmodium falciparum* for drug discovery and suppressing resistance. *Proc. Natl Acad. Sci. USA* **111**, 799–804 (2014).



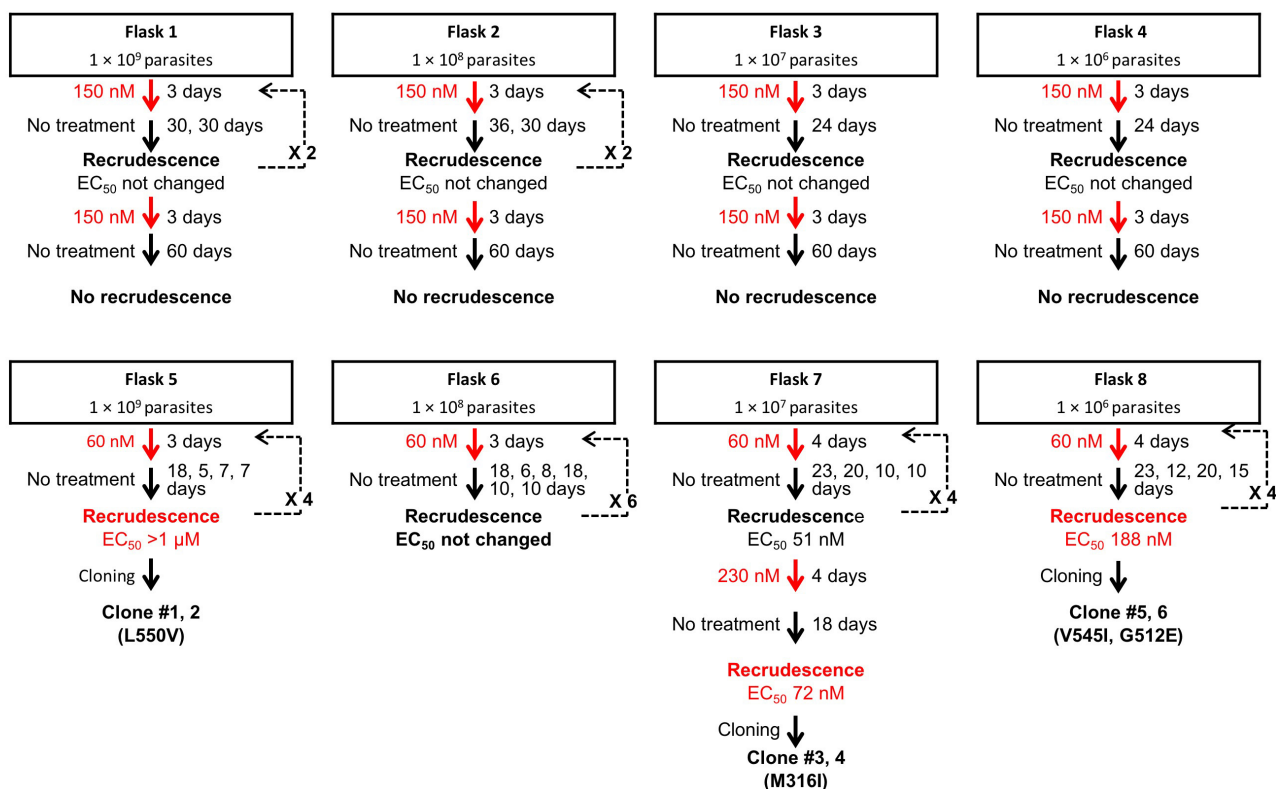
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Three screening-hit series yield new compound scaffolds against known targets.

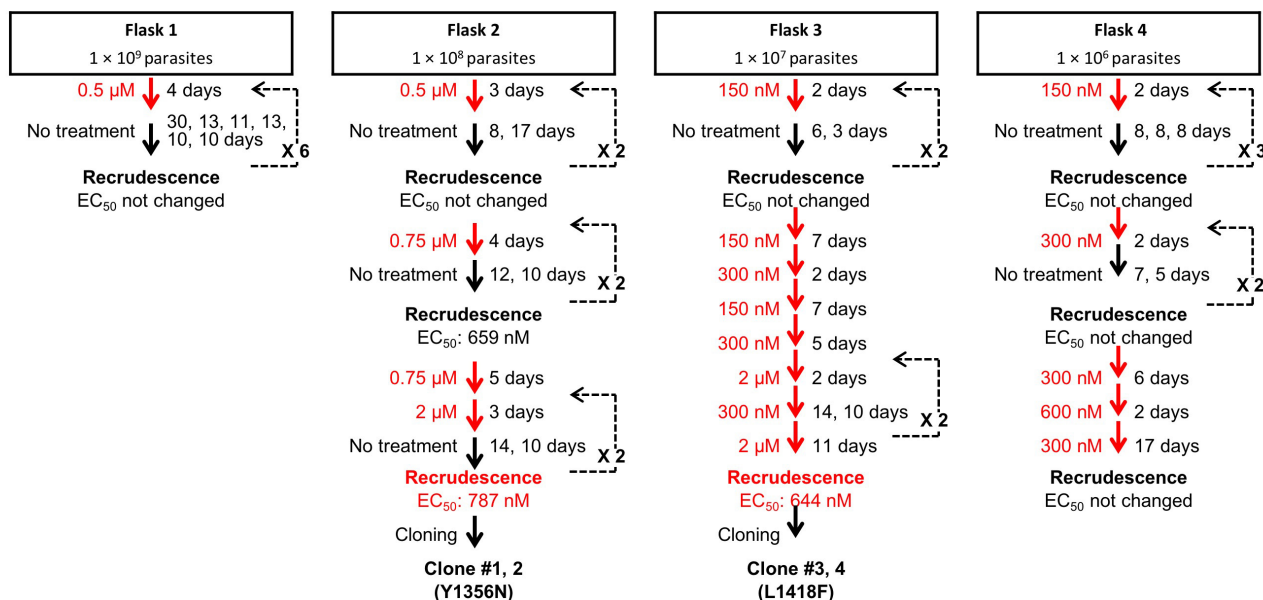
a–d, BRD0026 exhibits the same mode of action as NITD609 and showed moderate *in vitro* potency against asexual ($EC_{50} = 0.346 \mu M$) and late-sexual ($EC_{50} = 1.98 \mu M$) blood stages of the parasites and exhibited reduced potency against *P. falciparum* NITD609^R ($EC_{50} = 1.77 \mu M$), a transgenic strain carrying a point mutation in *P. falciparum* ATPase4 (ref. 9). *P. falciparum* ATPase4 is the presumed molecular target of NITD609 (ref. 9). **a, b**, Three of the eight possible stereoisomers (*R,S,R*; *S,S,S*; and *R,S,S*) of BRD0026 have activity. **c**, Initial characterization of BRD0026 showed good solubility in PBS and low cytotoxicity. **d**, Treatment with BRD0026 resulted in a rapid increase in the parasite cytosolic Na^+ concentration, while artesunate- or mefloquine-treated parasites maintained a constant cytosolic Na^+ concentration. This result suggests that parasites treated with BRD0026 are not able to counter the influx of Na^+ by actively extruding the cation, similar to the proposed mechanism for NITD609 (data are mean \pm s.d.; two biological and two technical replicates). **e–h**, BRD7539 targets and inhibits *P. falciparum* DHODH. BRD7539 showed excellent *in vitro* potency against liver-stages ($EC_{50} = 0.015 \mu M$) and asexual blood-stages ($EC_{50} = 0.010 \mu M$) of the parasite, conferring markedly reduced potency against *PfscDHODH*¹⁹. This strain heterologously expresses the cytosolic *S. cerevisiae* DHODH, which does not require ubiquinone as an electron acceptor. Thus, this transgenic strain is resistant to inhibitors of mitochondrial electron transport chain functions¹⁹. BRD7539 was tested against three different *P. falciparum* strains with mutations in mitochondrial genes targeted by other antimalarial agents: (i) TM90C6B strain, containing a point mutation in the quinol oxidase domain of *P. falciparum* cytochrome *b* (Q_o site) and resistant to atovaquone¹⁵; (ii) a *P. falciparum* CYTb^{G33V} mutant strain, selected against IDI5994 and containing a point mutation in the quinone reductase site of *P. falciparum* cytochrome *b* (Q_i site)¹⁵; and (iii) a *P. falciparum* DHODH^{E182D} mutant strain, selected against Genz-666136 and containing a point mutation in the *P. falciparum* DHODH gene⁷³. BRD7539 exhibited an approximately 59-fold shift in potency against the *P. falciparum* DHODH^{E182D} strain, whereas potency was unaffected in the TM90C6B and *P. falciparum* CYTb^{G33V} strains. BRD7539 also inhibits recombinant *P. falciparum* DHODH in an *in vitro*

biochemical assay ($IC_{50} = 0.033 \mu M$) but not the human orthologue. Altogether, these results demonstrate that BRD7539 targets *P. falciparum* DHODH. **e, f**, Only two (*S,S,S* and *R,S,S*) of eight possible stereoisomers of BRD7539 showed activity. **g**, *In vitro* growth inhibition assays showed no change in activity in *P. falciparum* CYTb^{G33V} and TM90C6B strains but exhibited a tenfold change in potency in *P. falciparum* DHODH^{E182D} strain, indicating that BRD7539 targets *P. falciparum* DHODH but not *P. falciparum* cytochrome *bc*₁. **h**, BRD7539 inhibited recombinant *P. falciparum* DHODH *in vitro* with an IC_{50} of 33 nM; no inhibition of the human orthologues was observed (data are mean \pm s.d. for two biological and two technical replicates). **i–m**, BRD73842 targets and inhibits *P. falciparum* PI4K. BRD73842 showed excellent *in vitro* activity against asexual ($EC_{50} = 0.069 \mu M$), late-sexual blood-stage ($EC_{50} = 0.643 \mu M$) and liver-stage ($EC_{50} = 0.459 \mu M$) parasites. **i, j**, The structure of BRD73842 indicates the required stereochemistry for activity (*R* stereoisomer). **k**, Initial characterization of BRD73842 showed good solubility and limited cytotoxicity. To gain insight into the mechanism of action of BRD73842, two resistant *P. falciparum* lines were evolved against BRD73842 from four independent cultures (a total of over 4×10^9 inocula, see Extended Data Fig. 2a). After more than 3 months of drug pressure, the EC_{50} values increased approximately 10- to 20-fold. Two clones were obtained from each culture. Sequence analyses revealed that all clones contain non-synonymous SNVs in PF3D7_0509800, the locus that encodes *P. falciparum* PI4K (Supplementary Table 3). **l**, To confirm that PI4K is the molecular target of BRD73842, the compound was assayed against purified recombinant *P. vivax* PI4K protein. BRD73842 selectively inhibits the kinase activity of *P. vivax* PI4K ($IC_{50} = 21$ nM), but not human PI4K. *P. falciparum* PI4K has been identified as the molecular target of two recently described antimalarial compounds, KAI407 (ref. 20) and MMV048 (ref. 21). (data are mean \pm s.d.; two biological and two technical replicates). **m**, The biphasic dose–response curve is a signature of *P. falciparum* PI4K inhibitors (data are mean \pm s.d.; three biological and three technical replicates). The EC_{50} values reported in this study are derived from the first transition of the dose–response curves (indicated by arrow).

a

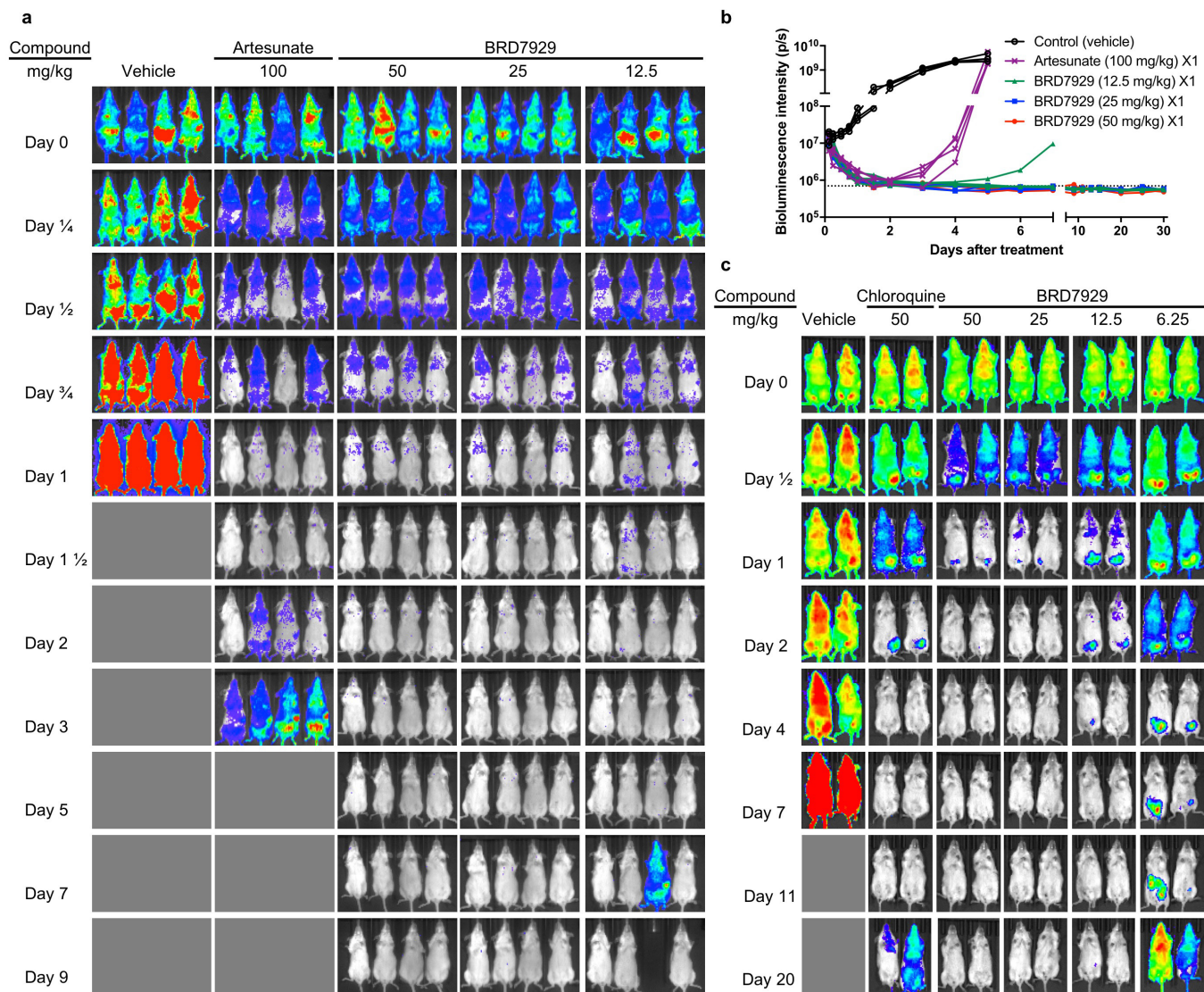


b



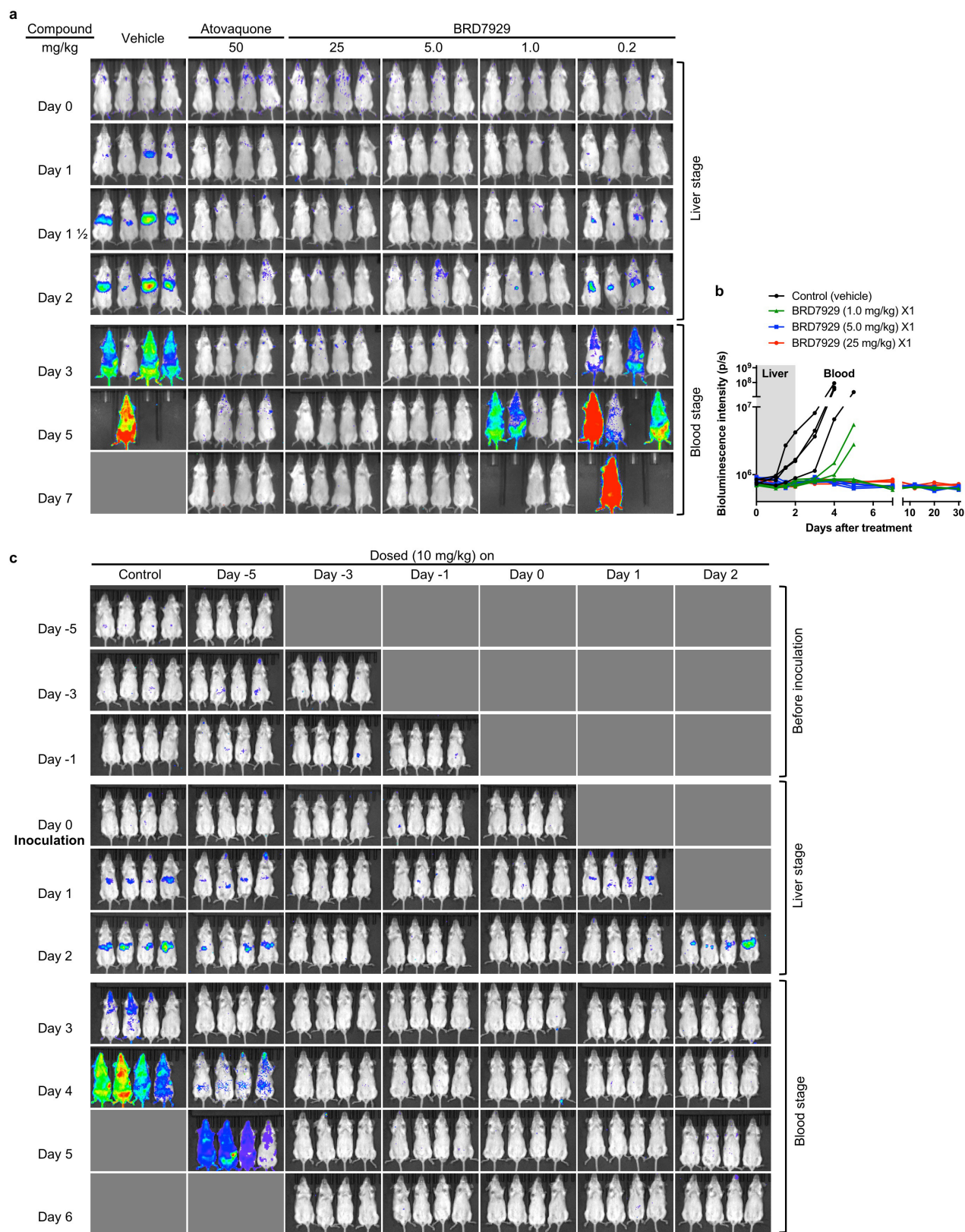
Extended Data Figure 2 | Resistance selection of BRD38427 and BRD1095. **a**, Over 3 months of intermittent and increasing resistance selection pressure of BRD73842 starting at 150 nM ($EC_{99.9}$) or 0.5 μM ($10 \times EC_{50}$) yielded two cultures showing a 13- to 16-fold EC_{50} shift. Two clonal lines from each culture were developed and subjected to

whole-genome sequencing. **b**, Over 3 months of intermittent pressure of BRD1095 at 60 nM ($EC_{99.9}$) or 150 nM ($10 \times EC_{50}$) yielded three cultures showing a 3- to 67-fold EC_{50} shift. Two clonal lines from each culture were developed and subjected to whole-genome sequencing.



Extended Data Figure 3 | *In vivo* blood-stage efficacy study of BRD7929. **a**, BRD7929 shows single-dose *in vivo* efficacy in a *P. berghei* model of malaria. CD-1 mice were inoculated intravenously with approximately 2×10^7 *P. berghei* (ANKA GFP-luc) blood-stage parasites intravenously 24 h before treatment and BRD7929 was administered as a single 50, 25, or 12.5 mg kg⁻¹ dose orally at 0 h ($n = 4$ for each group, this study was conducted once). Infections were monitored using IVIS. A single 100 mg kg⁻¹ dose of artesunate results in rapid suppression of parasites, but owing to its short half-life, the parasites re-emerge very quickly. A single 25 mg kg⁻¹ dose of BRD7929 resulted in 100% cure of the infected animals. One in four animals treated with a single oral dose of 12.5 mg kg⁻¹ showed recrudescence at 6 days after treatment, but all other animals administered with 12.5 mg kg⁻¹ were also completely parasite-free for 30 days. To ensure that no viable parasites remained, approximately 100 μ l of combined blood samples from the four animals treated with 25 mg kg⁻¹ of BRD7929 was intravenously injected into two naive mice and parasitaemia was monitored for an additional 30 days. No parasites were detected, suggesting that BRD7929 achieved a sterile cure for *P. berghei* with a single oral dose of as low as 25 mg kg⁻¹. The same colour scale is used for the all images; not all time-point

images are shown here. **b**, Bioluminescent intensity was quantified from each mouse and plotted against time. The dotted horizontal line represents the mean bioluminescence intensity level obtained from all the animals before the parasite inoculation. **c**, BRD7929 shows single-dose *in vivo* efficacy in a *P. falciparum* huRBC NSG mouse blood-stage model. huRBC NSG mice were inoculated intravenously with approximately 1×10^7 *P. falciparum* 3D7^{HLH/BRD} blood-stage parasites 48 h before treatment and BRD7929 was administered as a single 50, 25, 12.5 or 6.12 mg kg⁻¹ dose orally at 0 h ($n = 2$ for each group, this study was conducted once). Infections were monitored using the IVIS. No recrudescence was observed at doses as low as a single 12.5 mg kg⁻¹ of BRD7929 in the infected animals. To ensure that no viable parasites remained, approximately 350 μ l of combined blood samples from the two animals treated with 12.5 mg kg⁻¹ of BRD7929 was cultured *in vitro* and monitored for an additional 30 days. No parasites were detected, suggesting that BRD7929 achieved a sterile cure for *P. falciparum* 3D7^{HLH/BRD} with a single oral dose as low as 12.5 mg kg⁻¹ (see Fig. 4a). The same colour scale is used for the all images; not all time point images are shown. Images of mice treated with vehicle on days 11 and 20 are not shown, because the bioluminescent signal was too high to show in the same colour scale as other images.

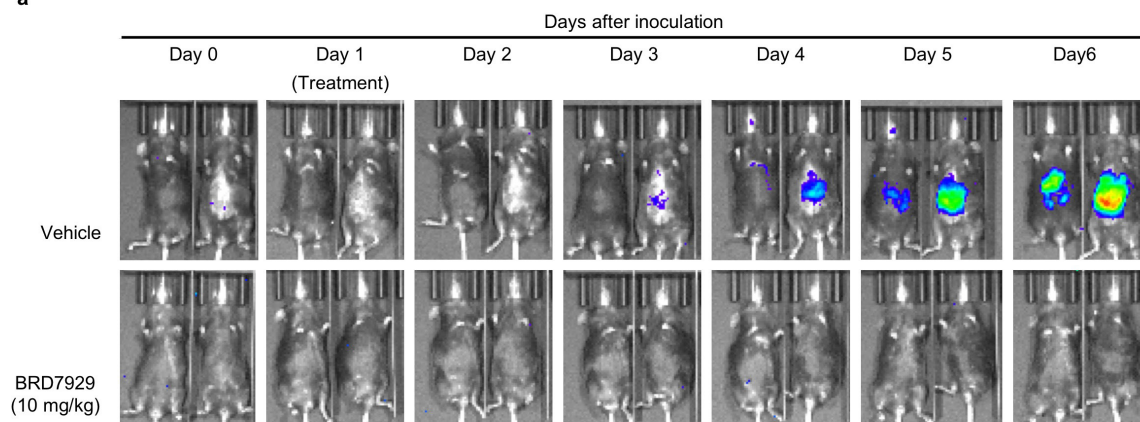


Extended Data Figure 4 | See next page for caption.

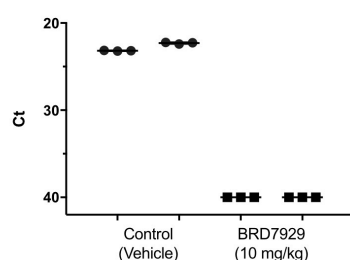
Extended Data Figure 4 | *In vivo* liver-stage efficacy study of BRD7929 in a mouse malaria model. **a**, BRD7929 shows single-dose causal prophylaxis in a *P. berghei* liver-stage model. CD-1 mice were inoculated intravenously with approximately 1×10^5 freshly dissected *P. berghei* ANKA luc-GFP sporozoites freshly dissected from *A. stephensi* salivary glands and immediately treated with a single oral dose of BRD7929 (25, 5, 1 or 0.2 mg kg^{-1}). Infections were monitored using IVIS; mice were monitored until day 30 to ensure complete cure. No recrudescence was observed at doses as low as a single 5 mg kg^{-1} of BRD7929 in the infected animals ($n = 4$ for each group, study conducted once). The same colour scale is used for the all images. Not all time point images are shown. **b**, Bioluminescent intensity was quantified from each mouse and plotted

against time. **c**, BRD7929 shows single-dose causal prophylaxis in a *P. berghei* liver-stage model up to 3 days before infection and two days after infection. CD-1 mice were infected with *P. berghei* and infections were monitored as described in **a**. Single oral doses of BRD7929 (10 mg kg^{-1}) were administered at days 5, 3, and 1 before infection (days -5 , -3 and -1), on day 0, and on days 1 and 2 after infection ($n = 4$ for each group, this study was conducted once). All dosing regimens except for the day -5 dose offered complete protection from infection for 32 days, indicating that BRD7929 has potent causal prophylaxis activity. The same colour scale is used for all images. Not all time-point images are shown.

a

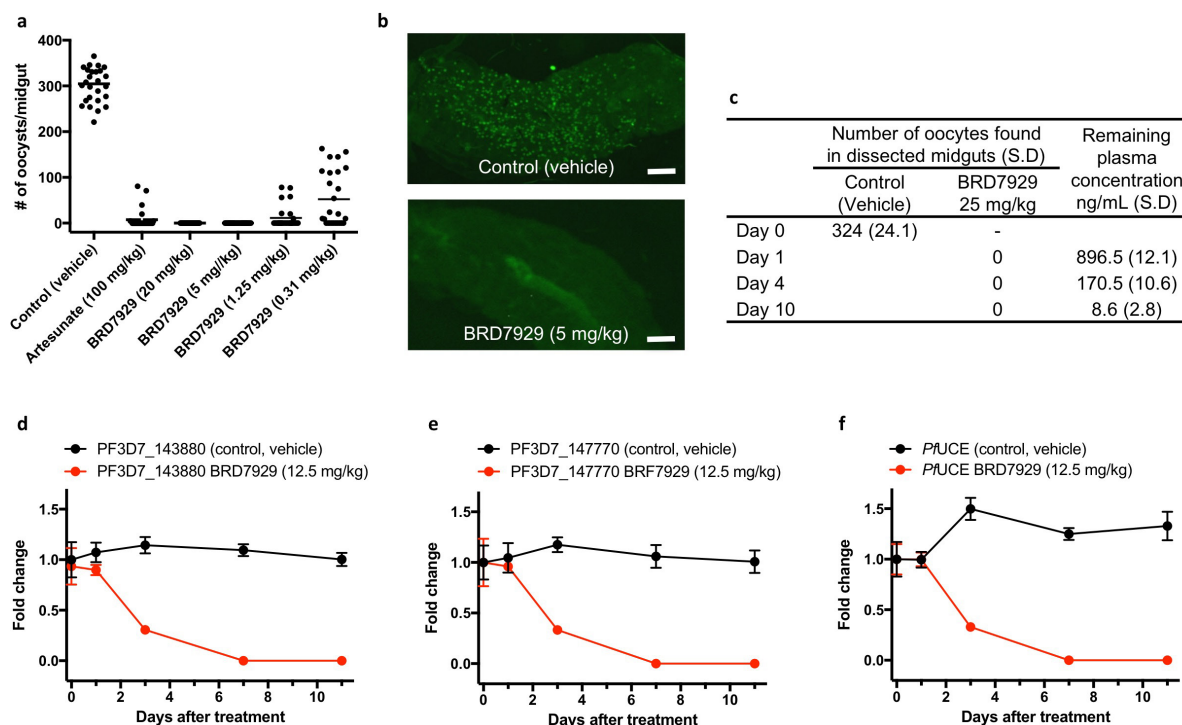


b



Extended Data Figure 5 | *In vivo* liver-stage efficacy study of BRD7929 in a humanized mouse model. **a**, BRD7929 shows single-dose *in vivo* efficacy in a *P. falciparum* huHep FRG-knockout mouse liver-stage model. huHep FRG knockout mice were inoculated intravenously with approximately 1×10^5 *P. falciparum* (NF54HT-GFP-luc) sporozoites and BRD7929 was administered as a single 10 mg kg^{-1} oral dose 1 day after inoculation ($n=2$ for each group, this study was conducted once). Infections were monitored using IVIS. The same colour scale is used for all images. No increase in bioluminescence intensity level was observed from the mice treated with BRD7929 (see Fig. 4b). **b**, Blood samples were also collected from each mouse 7 days after inoculation (the first day of the blood stage) and analysed for the presence of the blood-stage transcripts

PF3D7_0812600 (*P. falciparum* UCE) using qRT-PCR³² (two biological replicates for each group and three technical replicates for each biological sample). Each dot represents a technical replicate of a sample and each horizontal line represents a mean of technical replicates from each mouse. The presence of the blood-stage parasite specific transcripts was detected from the control (vehicle) mice, while no amplification of the marker was detected after 40 amplification cycles (C_t value = 40) from the mice treated with BRD7929. Primer efficiency and sensitivity of the primer pairs for *P. falciparum* UCE have a detection limit ranging between 10 and 100 transcript copies³³. Approximately 110 μl of combined blood samples from the two treated animals was also cultured *in vitro* and monitored for an additional 30 days but viable parasites were not detected.

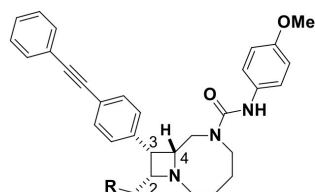


Extended Data Figure 6 | In vivo transmission-stage efficacy study of BRD7929. **a**, Oral doses of BRD7929 2 days before feeding mosquitoes upon infected mice resulted in complete blocking of transmission at 5 mg kg^{-1} , and reduced transmission activity at 1.25 mg kg^{-1} and 0.31 mg kg^{-1} ($n = 2$ for each group, this study was conducted once).

b, Mosquitoes fed on vehicle-treated mice showed heavy infection 1 week after feeding, while mosquitoes fed on treated mice showed no or very few oocysts in the midguts. Representative images are shown; scale bars, $100 \mu\text{m}$. **c**, To confirm that BRD7929 eliminates mature gametocytes in the host circulation rather than killing gametes, zygotes or ookinetes in the mosquito midgut, CD-1 mice infected with *P. berghei* (parasitaemia between 11 to 19%) were first treated with BRD7929 (oral, 25 mg kg^{-1}). Infected mice were then exposed to female *A. stephensi* mosquitoes for blood feeding 1, 4 or 10 days after the treatment. Blood samples were also obtained before the blood feedings to measure the plasma concentration of remaining BRD7929 ($n = 2$ for each group, this study was conducted

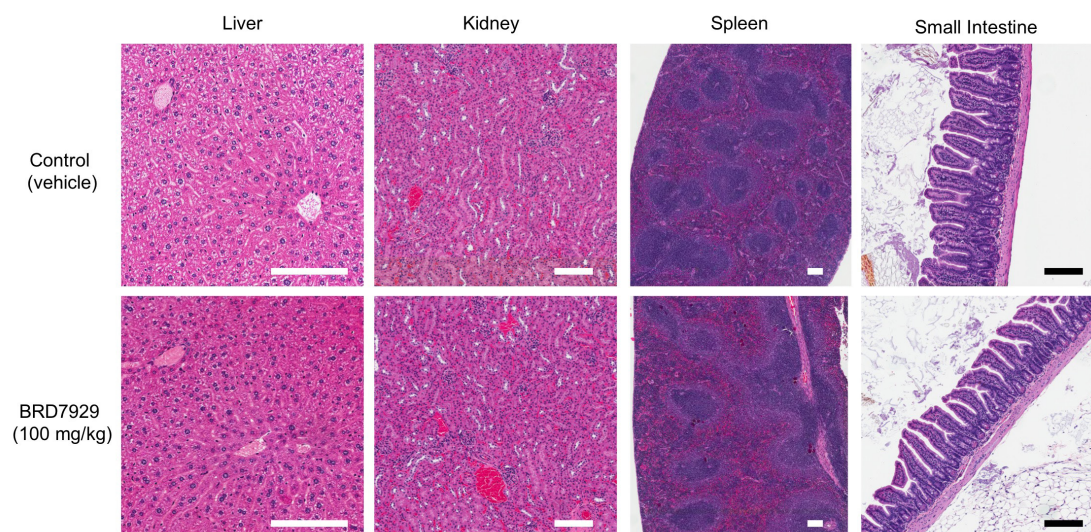
once). No oocysts were found in midguts dissected from mosquitoes from all time points, whereas 896.5 , 170.5 and 8.6 ng mL^{-1} of the compound remained in the circulation 1, 4 and 10 days after respectively treatment, respectively, suggesting that BRD7929 eliminated mature gametocytes in the mice. **d–f**, In vivo transmission-stage efficacy study of BRD7929 (humanized mouse model). huRBC NSG mice were infected with the blood-stage *P. falciparum* 3D7^{HLH/BRD} for 2 weeks to allow the gametocytes to mature fully and were treated with a single oral dose of BRD7929 (12.5 mg kg^{-1}). $n = 2$ for each group, this study was conducted once. Blood samples collected from vehicle- and BRD7929-treated mice were tested for the presence of gametocyte-specific transcripts using mature gametocyte marker (PF3D7_143880; **d**) and immature gametocyte marker (PF3D7_147770; **e**). PF3D7_1120200 (*P. falciparum* UCE), a constitutively expressed gene, was used as a positive control marker for parasite detection (**f**). Data are mean \pm s.d.; three technical replicates for each biological sample.

a

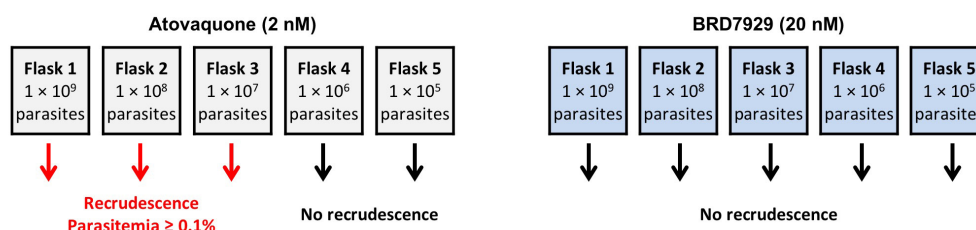


Compound	R=	BRD3444	BRD1095	BRD7929	BRD3316
		OH	NH ₂	NMe ₂	O(CH ₂) ₂ CO ₂ H
HepG2; CC ₅₀ (μM)		> 50	16	9	> 50
A549; CC ₅₀ (μM)		18	10	6	> 50
HEK 293; CC ₅₀ (μM)		45	16	10	> 50
Phototoxicity 3T3 NRU ^a *		Non-phototoxic		Non-phototoxic	
Reversible CYP inhibition [†] ; IC ₅₀ (μM)		> 10 (all)	4 (CYP1A)	> 10 (all)	> 10 (all)
Time-dependent CYP inhibition; k _{inact} /K _I (μM ⁻¹ L ⁻¹ min ⁻¹) [‡]		0.0158 (CYP3A)	negative (all)	negative (all)	negative (all)

b



c



d

# of inoculum	First day for recrudescence parasitemia to reach ≥ 0.1%	
	Atovaquone	BRD7929
1 × 10 ⁹	31	-
1 × 10 ⁸	37	-
1 × 10 ⁷	40	-
1 × 10 ⁶	-	-
1 × 10 ⁵	-	-

Extended Data Figure 7 | Safety and resistance propensity profiling of the bicyclic azetidine series. **a**, Results of *in vitro* cytotoxicity, phototoxicity and CYP inhibition assays. *Phototoxicity was assessed using the NIH 3T3 neutral red assay at Cyprotect; †CYP1A, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP3A; ‡CYP1A, CYP2C9, CYP2D6, CYP3A. **b**, Histopathology analysis of mice treated with a high dose (100 mg kg⁻¹) of BRD7929. CD-1 mice were orally treated with

100 mg kg⁻¹ BRD7929 and organs were collected 10 days after treatment. No significant tissue damage was detected. Representative images are shown here. Scale bars, 200 μm. **c**, **d**, Measurement of the minimal inoculum for resistance of BRD7929. Cultures containing various numbers of inoculum (1 × 10⁵–1 × 10⁹) were exposed to a constant level of drug pressure (EC₉₀). Parasites developed resistance to atovaquone at the lowest inoculum of 1 × 10⁷ but not to BRD7929.

Extended Data Table 1 | *In vitro* potency of BRD3444, BRD7929 and BRD3316 against multiple parasite stages

Species (strain)	Stage	EC ₅₀ (μM)		
		BRD3444	BRD7929	BRD3316
<i>P. falciparum</i> (Dd2)	Blood	0.009	0.005	0.019
<i>P. falciparum</i> (3D7 ^{HLH/BRD})	Blood		0.009	
<i>P. falciparum</i> (3D7)	Gametocyte (IV-V)	0.663	0.160	
<i>P. falciparum</i> (NF54)	Gametocyte (ID / D)*		0.270 / < 10	
<i>P. falciparum</i> (NF54)	Gametocyte (E / L) †	0.282 / 1.44		
<i>P. falciparum</i> (NF54)	Gamete formation (M / F) ‡	~1.00 / 0.804		
<i>P. falciparum</i> (NF54)	Liver	1.31	0.340	
<i>P. berghei</i> (ANKA)	Liver	0.140	0.162	
<i>P. cynomolgi</i> (M)	Liver (SF / LF) ¶	3.34 / 2.86	0.933 / 1.04	

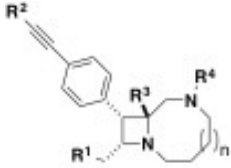
*Data indicate the results of a standard membrane-feeding assay⁷¹. Indirect (ID) exposure refers to parasites treated with varying drug concentrations for 24 h before mosquito feeding, while direct (D) refers to parasites treated with a single drug concentration (10 μM) immediately before blood feeding.

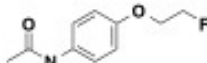
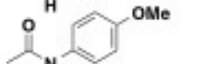


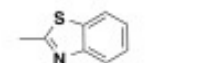

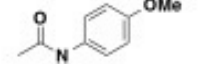

†Activity against early- (E, stages I–III) and late- (L, stages IV–V) stage gametocytes was assessed according to the protocol described previously⁵⁵.

‡Activity against male (M) and female (F) stage-V gametocytes was assessed in a dual gamete formation assay as described previously⁷¹. This assay (a standard membrane-feeding assay) is designed to determine the ability of compounds to either kill the mature *P. falciparum* male and female gametocytes directly or damage them in such a way that they cannot undergo onward development and form gametes in the mosquito midgut.

¶Activity against *P. cynomolgi* in primary rhesus hepatocytes was performed as described previously⁷². This assay measures inhibition of both the small form (SF, hypnozoite-like) and large form (LF, schizont) of intrahepatic *Plasmodium*.

Extended Data Table 2 | Structure–activity relationship study of the bicyclic azetidine series



	<i>Pf</i> Dd2 EC ₅₀ (μM)	<i>PfcPheRS</i> IC ₅₀ (μM)	R ¹	R ²	R ³	R ⁴	n
BRD8805	0.003	0.033	-NMe ₂	-Ph	-H		1
BRD7929	0.009	0.023	*	*	"		1
BRD1095	0.010	0.046	-NH ₂	*	"	*	1
BRD3444	0.011	0.033	-OH	*	"	*	1
BRD3316	0.022	0.029	-O(CH ₂) ₂ CO ₂ H	*	"	*	1
BRD4716	0.024	0.086	-NMe/Pr	*	"	*	1
BRD2132	0.048	0.179	-NMe(CH ₂) ₂ F	*	"	*	1
BRD0185	0.087	0.097	-OH	*	"	*	2
BRD8493	0.116	0.162		*	"	*	1
BRD6479	0.158	0.233		*	"	*	1
BRD4873	0.261	0.221	-OH	-2-CNPh	"	*	1
BRD9599	0.850	0.366	*	-Ph	"	*	0
BRD2936	1.87	29.4	*	*	-CH ₂ OH	*	1
BRD5349	8.32	30.9	*	*	-H		1
BRD5774	12.2	23.4	*		"		1
BRD8260	19.5	34.6	*	-Ph	"		1

The structures of 16 bicyclic azetidine analogues with varying potency against asexual blood-stage parasites (Dd2), along with their corresponding inhibition of the *P. falciparum* PheRS activity in a biochemical assay. Aminoacylation inhibition activities were characterized using purified recombinant PheRS in which a range of inhibitor concentrations was used to determine IC₅₀ values. The biochemically derived IC₅₀ values correlate extremely well ($r^2 = 0.89$) with the EC₅₀ determined using the blood-stage parasite growth-inhibition assay (see Fig. 3d).

Extended Data Table 3 | *In vitro* and *in vivo* pharmacokinetic properties of the bicyclic azetidine series

	BRD3444*		BRD1095*		BRD7929*			BRD7929†		BRD3316*	
<i>Pf</i> , Dd2 EC ₅₀ (nM)	9		10		9					23	
PBS solubility (μM)	< 1		25		15					91	
Mouse Plasma protein binding (%)	99.9		99.3		99.9						
Mouse Cl _{int} (μL/min/mg)	248		< 20		21					38	
Human Cl _{int} (μL/min/mg)	142		< 20		31					34	
HepG2 CC ₅₀ (μM)	> 50		15.6		9					> 50	
hERG IC ₅₀ (μM)	5.2		5.1		2.1					> 10	
Route (mg/kg)	IV (3)	PO (10)	IV (3)	PO (10)	IV (2.5)	IV (2.5)‡	PO (10)	PO (3)	PO (9)	IV (3.2)	PO (13)
C _{max} (μM)		0.6		0.6			0.54	0.21	0.6		6.8
T _{max} (hr)		0.5		4			8	12	12		1
T _{1/2} (hr)	3.7	3.2		28.8	N.C	32				2.3	2.4
AUC _{0-t} (μM*hr)	1.2¶	4¶	7¶	11.7¶	3.5¶	9#	11¶	6.4¶	19.7¶	13.2¶	33.5¶
AUC _{0-inf} (μM*hr)	1.4	4	14.9			11.2		7.2	22.6	13.2	33.5
MRT _{0-inf} (hr)	2.8		39.2		40.5	45		35.4	37.8	3.3	3.9
V _{ss} (L/kg)	12		16		24	19				1.4	
F (%)	86		50				79.5§				63
CL (mL/min/kg)	72		6.7		9.9	7.1				7.1	

BRD3444 and BRD1095 were formulated in 70% PEG400 and 30% aqueous glucose (5% in H₂O) for intravenous and oral dosing and pharmacokinetics were determined in CD-1 mice as described in Methods. Pharmacokinetic studies of BRD3444 and BRD1095 were performed by ChemPartner Co., Ltd and were estimated by a non-compartmental model using WinNonlin 6.2. BRD7929 and BRD3316 were formulated in 10% ethanol, 4% Tween, 86% saline for both intravenous and oral dosing. Pharmacokinetics in *P. falciparum* 3D7^{HLH/BRD}-infected NSG mice was determined on dried blood spot samples from infected NSG mice using standard methods. Pharmacokinetics parameters for BRD7929 and BRD3316 were estimated by a non-compartmental model using proprietary Eisai software. Cl_{int}, intrinsic clearance; CL, clearance; MRT, mean residence time; N.C, not calculated owing to insufficient data; V_{ss}, steady-state volume of distribution.

*Pharmacokinetic in CD-1 mice.

†Pharmacokinetic in *P. falciparum* 3D7^{HLH/BRD}-infected NSG mice.

‡Intravenously determined in a separate assay over 72 h to determine half-life.

¶t = 24 h.

#t = 72 h.

§Per cent value based on initial intravenous study at 24 h.

Frizzled proteins are colonic epithelial receptors for *C. difficile* toxin B

Liang Tao^{1,2}, Jie Zhang^{1,2*}, Paul Meraner^{3*}, Alessio Tovaglieri⁴, Xiaoqian Wu⁵, Ralf Gerhard⁶, Xinjun Zhang^{7,8}, William B. Stallcup⁹, Ji Miao^{4,10}, Xi He^{7,8}, Julian G. Hurdle⁵, David T. Breault^{4,10,11}, Abraham L. Brass^{3,12} & Min Dong^{1,2}

***Clostridium difficile* toxin B (TcdB) is a critical virulence factor that causes diseases associated with *C. difficile* infection. Here we carried out CRISPR–Cas9-mediated genome-wide screens and identified the members of the Wnt receptor frizzled family (FZDs) as TcdB receptors. TcdB binds to the conserved Wnt-binding site known as the cysteine-rich domain (CRD), with the highest affinity towards FZD1, 2 and 7. TcdB competes with Wnt for binding to FZDs, and its binding blocks Wnt signalling. FZD1/2/7 triple-knockout cells are highly resistant to TcdB, and recombinant FZD2–CRD prevented TcdB binding to the colonic epithelium. Colonic organoids cultured from FZD7-knockout mice, combined with knockdown of FZD1 and 2, showed increased resistance to TcdB. The colonic epithelium in FZD7-knockout mice was less susceptible to TcdB-induced tissue damage *in vivo*. These findings establish FZDs as physiologically relevant receptors for TcdB in the colonic epithelium.**

Infection of the colon by the opportunistic Gram-positive bacterium *C. difficile* leads to a range of manifestations from diarrhoea to life-threatening pseudomembranous colitis and toxic megacolon^{1–5}. It is the most common cause of antibiotic-associated diarrhoea and a leading cause of gastroenteritis-associated death in developed countries, accounting for nearly half a million cases and 29,000 deaths annually in the United States⁶. Two homologous exotoxins, TcdA and TcdB, are the causal agents for diseases associated with *C. difficile* infection^{4,7–9}. These toxins enter cells via receptor-mediated endocytosis and inactivate small GTPases by glucosylating a key residue, resulting in cell rounding and eventual cell death^{4,7,10}. Of the two toxins, TcdB alone is capable of causing the full spectrum of diseases, as TcdA[–]B⁺ strains have been clinically isolated and engineered TcdA[–]B⁺ strains induced death in animal models^{11–14}.

How TcdB targets the colonic epithelium remains unknown. TcdB can enter a variety of cell lines, suggesting that its receptor(s) are widely expressed in transformed cells. It has also been reported that TcdB is enriched in the heart after injection into zebrafish embryos¹⁵. Chondroitin sulfate proteoglycan 4 (CSPG4, also known as neuronal antigen 2 (NG2)) has been identified as a TcdB receptor in a short hairpin RNA (shRNA)-mediated knockdown screen¹⁶, and was shown to be a functional receptor for TcdB in HeLa cells and in HT-29 cells, a human colorectal cell line. However, CSPG4 is not expressed in the colonic epithelium¹⁷. Poliovirus receptor-like 3 (PVRL3; also known as nectin-3) was recently identified from a gene-trap insertional mutagenesis screen in Caco-2 cells, a human colorectal cell line, as a factor involved in necrotic cell death (cytotoxicity) induced by TcdB¹⁸, but whether it functions as a TcdB receptor remains to be established.

Here we carried out unbiased genome-wide screens using the CRISPR–Cas9 approach^{19,20} and identified the FZDs as TcdB receptors. Using colonic organoid models and FZD7-knockout mice, we

established FZDs as physiologically relevant receptors for TcdB in the colonic epithelium.

CRISPR–Cas9 screen for TcdB receptors

The C-terminal domains of TcdA and TcdB contain a region known as combined repetitive oligopeptides (CROPs) (Extended Data Fig. 1a), which can bind carbohydrates and may mediate toxin binding to cells²¹. Recent studies suggested the presence of an additional receptor-binding region beyond the CROPs^{22–25}. Consistently, we found that a truncated toxin (TcdB_{1–1830}) lacking the CROPs induced cell rounding in various cell lines at picomolar concentrations (Extended Data Fig. 1b–d)²⁶. To identify both the receptor(s) recognized by the CROPs and the receptor(s) recognized by other regions, we carried out two separate screens, with either full-length TcdB or TcdB_{1–1830} (Fig. 1a).

HeLa cells that stably express RNA-guided endonuclease Cas9 were transduced with lentiviral libraries that express short guide RNAs (sgRNA) targeting 19,052 genes, with six sgRNAs per gene¹⁹. After four rounds of selection with increasing concentrations of toxins, the sgRNA sequences from the surviving cells were identified via next-generation sequencing (NGS). We ranked candidate genes based on the number of unique sgRNAs versus NGS reads (Fig. 1b, c, Extended Data Fig. 2 and Source Data).

UDP-glucose pyrophosphorylase (UGP2) stood out in both screens (Fig. 1b, c). UGP2 is a cytosolic enzyme producing UDP-glucose, which is used by TcdA and TcdB to glucosylate small GTPases. Mutations in UGP2 have been shown to render cells resistant to TcdA and TcdB^{27,28}. Besides UGP2, the top hit from the full-length TcdB screen is CSPG4 (Fig. 1b), confirming a previous report that identified CSPG4 as a TcdB receptor¹⁶. The highest-ranking plasma membrane protein from the TcdB_{1–1830} screen is FZD2 (Fig. 1c). FZD2 is a member of the FZD family of receptors for Wnt signalling, which is a key

¹Department of Urology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Microbiology and Immunobiology and Department of Surgery, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Microbiology and Physiological Systems (MaPS), University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA. ⁴Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ⁵Center for Infectious and Inflammatory Diseases, Texas A & M Health Science Center, Houston, Texas 77030, USA. ⁶Institute of Toxicology, Hannover Medical School, 30625 Hannover, Germany. ⁷The F. M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Department of Neurology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁹Tumor Microenvironment and Cancer Immunology Program, Sanford-Burnham Prebys Medical Discovery Institute, Cancer Center, La Jolla, California 92037, USA. ¹⁰Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. ¹²Gastroenterology Division, Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA.

*These authors contributed equally to this work.

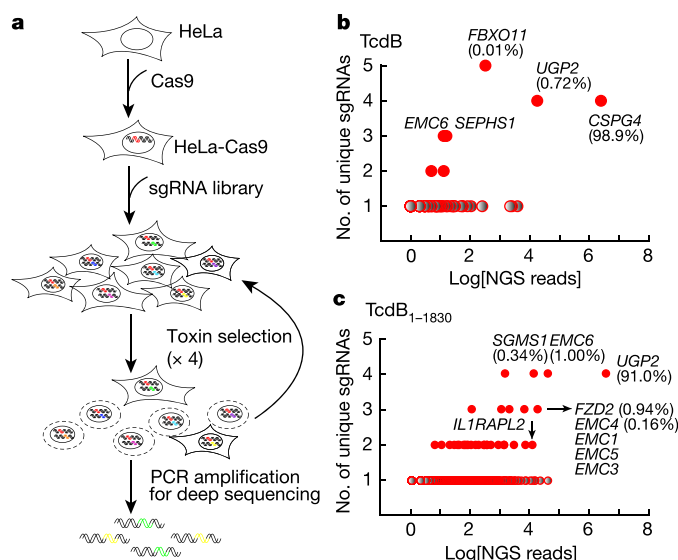


Figure 1 | Genome-wide CRISPR-Cas9-mediated screens to identify host factors for TcdB. **a**, Schematic drawing of the screen process. PCR, polymerase chain reaction. **b**, **c**, Genes identified in the screens with TcdB (**b**) or TcdB₁₋₁₈₃₀ (**c**). The y axis is the number of unique sgRNAs for each gene. The x axis represents the number of sgRNA reads for each gene. The percentages of the sgRNA reads of top-ranking genes among total sgRNA reads are noted.

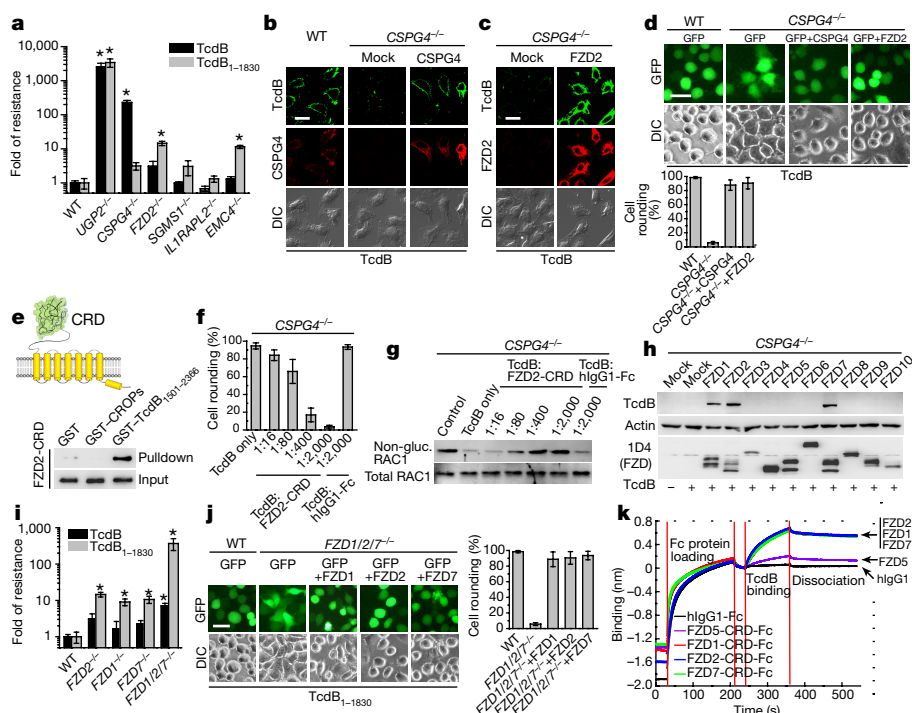


Figure 2 | FZDs are functional receptors for TcdB. **a**, The sensitivities of the indicated HeLa knockout cells to TcdB and TcdB₁₋₁₈₃₀ were quantified using the cytopathic cell-rounding assay (see Extended Data Fig. 3) and normalized to wild-type (WT) HeLa cells as fold of resistance. The experiments were repeated three times. **b**, **c**, Immunostaining analysis showed that TcdB binding (10 nM, 10 min) to CSPG4^{-/-} cells was reduced (**b**). Ectopic expression of rat CSPG4 increased binding of TcdB. Transfection of FZD2 also increased TcdB binding to CSPG4^{-/-} cells (**c**). Scale bar, 20 μm. **d**, Ectopic expression of CSPG4 or FZD2 restored TcdB entry into CSPG4^{-/-} cells, resulting in cell rounding (5 pM, 3 h). Green fluorescent protein (GFP)-marked transfected cells. Scale bar, 50 μm. **e**, A schematic illustration of FZD (top). Fc-tagged FZD2-CRD binds to GST-tagged TcdB₁₅₀₁₋₂₃₆₆, but not to GST-tagged CROPs. **f**, **g**, FZD2-CRD prevented TcdB (300 pM, 3 h)

signalling pathway regulating proliferation and self-renewal of colonic epithelial cells^{29,30}. Besides FZD2, an unusual group of high-ranking hits are subunits of the endoplasmic reticulum membrane protein complex (EMC)^{31,32}.

To validate the screening results, we generated UGP2^{-/-}, CSPG4^{-/-}, FZD2^{-/-} and EMC4^{-/-} HeLa cell lines using the CRISPR-Cas9 approach (Supplementary Table 1). Two additional knockout cells were also generated and examined: sphingomyelin synthase 1 (SGMS1^{-/-}) and interleukin-1 receptor accessory protein-like 2 (IL1RAPL2^{-/-}) (Fig. 1c). These cells were challenged with either TcdB or TcdB₁₋₁₈₃₀ using the cytopathic cell-rounding assay¹ (Extended Data Fig. 3a, b). UGP2^{-/-} cells were highly resistant (~3,000-fold) to both TcdB and TcdB₁₋₁₈₃₀ compared with wild-type HeLa cells (Fig. 2a and Supplementary Table 2). CSPG4^{-/-} cells showed increased resistance to TcdB (~240-fold), but not to TcdB₁₋₁₈₃₀. FZD2^{-/-} and EMC4^{-/-} cells both showed increased resistance (~15 and ~11-fold, respectively) to TcdB₁₋₁₈₃₀, but not to TcdB. SGMS1^{-/-} and IL1RAPL2^{-/-} cells did not show significant changes in sensitivity to toxins under our assay conditions. Increased resistance of UGP2^{-/-}, CSPG4^{-/-}, FZD2^{-/-} and EMC4^{-/-} cells was further confirmed by immunoblot analysis for glucosylation of RAC1, a small GTPase (Extended Data Fig. 3c).

CSPG4 is a CROP-dependent TcdB receptor

We next focused on CSPG4 and FZD2 as potential TcdB receptors. Binding of TcdB to CSPG4^{-/-} cells was reduced compared with wild-type HeLa cells (Fig. 2b). Ectopic expression of rat CSPG4 restored

entry into CSPG4^{-/-} cells, measured by the cell-rounding assay (**f**) and glucosylation (gluc.) of RAC1 (**g**). Human IgG1-Fc (hlgG1-Fc) is a control. **h**, Transfection of either FZD1, 2 or 7 increased TcdB binding (10 nM, 10 min) to CSPG4^{-/-} cells, assayed by immunoblot analysis of cell lysates. Actin is a loading control. **i**, The sensitivities of FZD1^{-/-}, FZD2^{-/-}, FZD7^{-/-} and FZD1/2/7^{-/-} cells to TcdB and TcdB₁₋₁₈₃₀ were analysed as described in **a**. **j**, Ectopic expression of FZD1, 2 or 7 restored TcdB₁₋₁₈₃₀ entry into FZD1/2/7^{-/-} cells (300 pM, 3 h). Scale bar, 50 μm. **k**, Characterization of TcdB binding to Fc-tagged CRDs of FZD1, 2, 5 and 7 using the BLI assay (see Supplementary Table 3 for K_d analysis). Representative images are from one of three independent experiments. Error bars indicate mean ± standard deviation (s.d.), *n* = 6, **P* < 0.005, *t*-test.

binding and entry of TcdB (Fig. 2b, d). TcdB binds directly to purified rat CSPG4 extracellular domain fragments (CSPG4-EC) independent of the chondroitin sulfate glycan in CSPG4 (Extended Data Fig. 4a, b). These results are consistent with a previous report¹⁶. However, contrary to the previous suggestion that CSPG4 does not bind to the CROPs¹⁶, we conclude that the CROPs are essential for TcdB binding to CSPG4 because: (1) TcdB_{1–1830} does not bind to either purified CSPG4-EC or CSPG4 expressed on cell surfaces (Extended Data Fig. 4b, c); (2) *CSPG4*^{−/−} cells showed similar levels of sensitivity to TcdB_{1–1830} as wild-type cells (Fig. 2a); and (3) the CROPs are capable of competing with TcdB for binding to CSPG4 on cell surfaces (Extended Data Fig. 4d, e). We note that the previous study used TcdB_{1851–2366} as the CROPs¹⁶. Recent structural studies confirmed that the CROP region starts around residue 1834 instead of 1851 (ref. 33). Here we used full-length CROPs (residues 1831–2366). It is possible that the 1831–1850 region is required for TcdB binding to CSPG4.

FZDs are CROP-independent receptors

Transfecting *CSPG4*^{−/−} cells with FZD2 also increased binding of TcdB (Fig. 2c) and restored entry of TcdB into *CSPG4*^{−/−} cells (Fig. 2d), suggesting that FZD2 is an alternative receptor. In contrast to CSPG4, ectopically expressed FZD2 increased binding of TcdB_{1–1830} and TcdB_{1501–2366} on cell surfaces, but not the CROPs (Extended Data Fig. 4c, f), suggesting that it is a CROP-independent receptor. FZD2 is a seven-pass transmembrane protein and contains a sole distinct extracellular domain known as the CRD (Fig. 2e)²⁹. Consistently, recombinant Fc-tagged FZD2-CRD binds directly to glutathione S-transferase (GST)-tagged TcdB_{1501–2366}, but not to the CROPs in pulldown assays (Fig. 2e).

It is possible that CSPG4 is expressed at a much higher level than FZD2 in HeLa cells, which may explain why TcdB binding to *CSPG4*^{−/−} cells is barely detectable using immunostaining and immunoblot assays. Notably, TcdB can enter *CSPG4*^{−/−} cells at picomolar concentrations, as detected by the sensitive cytopathic cell-rounding assay (Fig. 2f). Such entry is blocked by recombinant FZD2-CRD in a dose-dependent manner, as evidenced by a lack of cell rounding and RAC1 glucosylation (Fig. 2f, g), suggesting that endogenous FZD2 mediates TcdB binding and entry in *CSPG4*^{−/−} cells.

The FZD family includes ten members (FZD1–10) in humans²⁹. The ectopic expression of FZD1, 2 and 7 each increased binding of TcdB to *CSPG4*^{−/−} cells (Fig. 2h and Extended Data Fig. 5a), probably because the CRDs of FZD1, 2 and 7 share ~98% sequence similarity (Extended Data Fig. 5b)²⁹. Consistently, FZD7-CRD, but not FZD8-CRD, when expressed on cell surfaces via a fused glycoposphatidylinositol (GPI) anchor, mediated strong binding of TcdB to cells (Extended Data Fig. 5c).

HeLa cells express multiple FZDs³⁴. We next generated FZD1 and FZD7 single-knockout HeLa cells, as well as FZD1/2/7 triple-knockout cells. *FZD1/2/7*^{−/−} cells exhibited normal growth rates, probably because HeLa cells still express other FZDs. *FZD1*^{−/−} and *FZD7*^{−/−} cells showed reductions in sensitivity to TcdB_{1–1830} similar to those of *FZD2*^{−/−} cells (Fig. 2i). *FZD1/2/7*^{−/−} cells were highly resistant to TcdB_{1–1830} (~300-fold), confirming that FZD1, 2 and 7 all contribute to TcdB_{1–1830} entry into HeLa cells. Transfection of FZD1, 2 or 7 restored TcdB_{1–1830} entry into *FZD1/2/7*^{−/−} cells (Fig. 2j). *FZD1/2/7*^{−/−} cells also become ~10-fold more resistant to full-length TcdB than wild-type cells (Fig. 2i), indicating that endogenous FZD1, 2 and 7 are responsible for a portion of TcdB entry into wild-type HeLa cells. *FZD1/2/7*^{−/−} cells showed the same level of sensitivity to TcdA as wild-type cells (Extended Data Fig. 5d), confirming that the resistance of *FZD1/2/7*^{−/−} cells is specific to TcdB.

We further quantified the binding kinetics between CRDs of FZD1, 2, and 7 and TcdB using the bio-layer interferometry (BLI) assay. The results revealed a single binding site with low nanomolar affinities (dissociation constant (K_d) = 32 nM for FZD1, 19 nM for FZD2, and 21 nM for FZD7) (Fig. 2k, Extended Data Fig. 5e and Supplementary

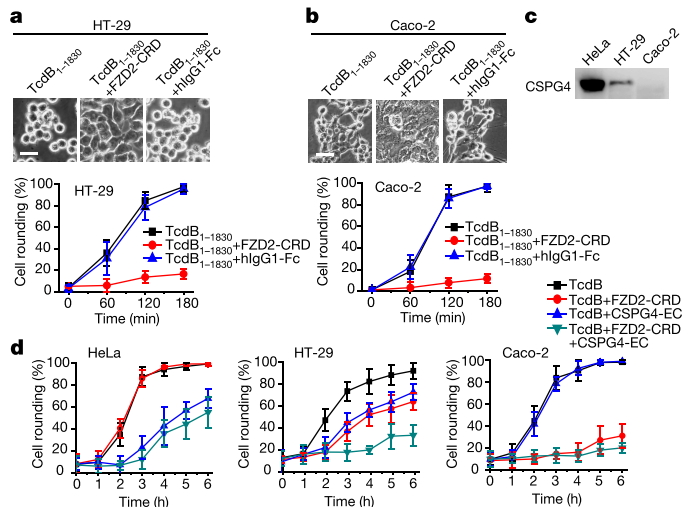


Figure 3 | FZDs versus CSPG4 in cell lines. **a, b**, FZD2-CRD protected HT-29 (**a**) and Caco-2 cells (**b**) from TcdB_{1–1830} (300 pM, 3 h).

Representative images are from one of three independent experiments. Scale bars: 25 μ m (**a**) or 50 μ m (**b**). **c**, Expression of CSPG4 in HeLa, HT-29 and Caco-2 cells was examined via immunoblot analysis of cell lysates. One experiment from four is shown. **d**, Protection from TcdB using FZD2-CRD and CSPG4-EC on HeLa (5 pM TcdB), HT-29 (50 pM TcdB) and Caco-2 (150 pM TcdB) cells was quantified by the cytopathic cell-rounding assay. Representative images are shown in Extended Data Fig. 6b. Error bars indicate mean \pm s.d.

Table 3). Furthermore, FZD2-CRD showed the same binding affinity to TcdB_{1–1830} (K_d = 17 nM) as to full-length TcdB (Extended Data Fig. 5f). FZD5-CRD also binds to TcdB when measured by the sensitive BLI assay, but with a much weaker affinity than FZD1, 2 and 7 (K_d = 670 nM) (Fig. 2k and Extended Data Fig. 5e). It is possible that additional FZD family members may function as low-affinity receptors for TcdB.

The finding that *EMC4*^{−/−} cells showed a similar level of toxin resistance as *FZD2*^{−/−} cells is also consistent with FZDs being TcdB receptors (Fig. 2a). Although its function remains to be established, the EMC appears to be critical for the folding/stability of multi-transmembrane proteins^{35,36}. Consistently, expression of transfected FZD1, 2 or 7 was reduced in *EMC4*^{−/−} cells compared with wild-type cells (Extended Data Fig. 5g, h).

CSPG4 versus FZDs in cell lines

We next addressed whether TcdB is capable of simultaneous binding to both CSPG4 and FZDs. As shown in Extended Data Fig. 6a, FZD2-CRD binds to TcdB pre-bound by immobilized CSPG4-EC on the microtitre plate, confirming that CSPG4 and FZDs do not compete with each other for binding to TcdB.

We then examined the receptors responsible for TcdB entry in HT-29 and Caco-2 cells, which are known to express multiple FZDs³⁷. FZD2-CRD protected both HT-29 and Caco-2 from TcdB_{1–1830} (Fig. 3a, b), suggesting that FZDs are functional receptors in these two cell lines. Interestingly, CSPG4 is expressed at high levels in HeLa, at much lower levels in HT-29, and is undetectable in Caco-2 cells (Fig. 3c). Consistently, CSPG4-EC alone was sufficient to reduce TcdB entry into HeLa cells, whereas a combination of CSPG4-EC and FZD2-CRD was required to reduce TcdB entry into HT-29 cells, and FZD2-CRD alone protected Caco-2 cells (Fig. 3d and Extended Data Fig. 6b). These data suggest that CSPG4 and FZDs represent non-competing TcdB receptors, each capable of mediating binding and entry of TcdB. Their particular contribution in a given cell type may depend on their expression levels.

We also tested the potential role of PVRL3. Ectopically expressed PVRL3 did not increase either binding or entry of TcdB into

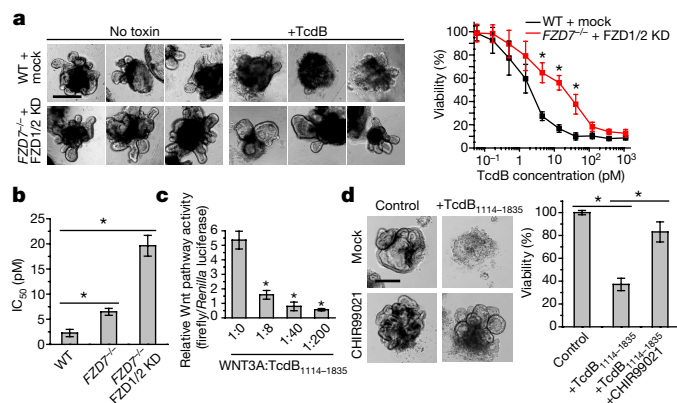


Figure 4 | FZDs are receptors for TcdB in colonic organoids.

a, Left, three sets of representative DIC images of wild-type (WT) and *FZD7*^{-/-} plus *FZD1/2*-knockdown (KD) organoids exposed to TcdB (0.5 pM, 3 days). Right, viability of organoids exposed to TcdB for 3 days was quantified by the MTT assay. *n* = 6, **P* < 0.005, *t*-test. **b**, The half-maximum inhibitory concentration (IC₅₀; the TcdB concentration that results in 50% viability after 3 days) of wild-type, *FZD7*^{-/-} and *FZD7*^{-/-} plus *FZD1/2*-knockdown organoids were quantified as described in **a**. *n* = 8, **P* < 0.005, one-way analysis of variance (ANOVA). **c**, TcdB₁₁₁₄₋₁₈₃₅ blocked WNT3A-mediated signalling in 293T cells. *n* = 6, **P* < 0.005, *t*-test. **d**, Viability of colonic organoids after exposure to TcdB₁₁₁₄₋₁₈₃₅ (25 nM), with or without CHIR99021 (5 μM), was quantified by the MTT assay. *n* = 8, **P* < 0.005, one-way ANOVA. Scale bars, 200 μm. Representative images are from one of three independent experiments. Error bars indicate mean ± s.d.

CSPG4^{-/-} HeLa cells (Extended Data Fig. 7a, b). The recombinant ecto-domain of PVRL3 failed to protect Caco-2 cells from TcdB in the cytopathic cell-rounding assay, whereas FZD2-CRD protected cells (Extended Data Fig. 7c). Thus, PVRL3 is probably not a relevant receptor for the cytopathic effect of TcdB in HeLa and Caco-2 cells.

FZDs are TcdB receptors in colonic organoids

To determine the receptors that mediate TcdB entry into the colonic epithelium, we first used colonic organoids, an *in vitro* 'mini-gut' model that recapitulates many important features of normal colonic epithelium³⁸. Exposure to TcdB caused dose-dependent atrophy and death of organoids, which was quantified using a viability assay (Fig. 4a). We found that TcdB₁₋₁₈₃₀ and TcdB were equally potent, suggesting that CSPG4 is not a relevant receptor in colonic organoids (Extended Data Fig. 8a). It has been reported that CSPG4 is not expressed in the colonic epithelium¹⁷, which was confirmed by immunoblot analysis of colonic organoids and isolated mouse colonic epithelium (Extended Data Fig. 8b).

We next used colonic organoids cultured from *FZD7*-knockout mice, combined with adenovirus-mediated knockdown of *FZD1* and *FZD2* (Extended Data Fig. 8c, d). *FZD7* is critical for maintaining intestinal organoids, but *FZD7*^{-/-} organoids can be cultured in the presence of CHIR99021, a small-molecule inhibitor of glycogen synthase kinase-3 (GSK3), which activates Wnt/β-catenin signalling downstream of FZDs³⁹. *FZD7*^{-/-} organoids showed threefold more resistance to TcdB than wild-type organoids (Fig. 4b). Further knockdown of *FZD1/2* in *FZD7*^{-/-} organoids yielded ninefold greater resistance to TcdB than wild-type organoids (Fig. 4b), demonstrating that FZDs are relevant TcdB receptors in colonic organoids.

As both TcdB and Wnt bind to the FZD-CRD, we examined whether TcdB binding competes with Wnt and inhibits Wnt signalling. We used a non-toxic TcdB fragment (residues 1114–1835), which contains the FZD-binding region but not the enzymatic domain of TcdB. This fragment blocked WNT3A-mediated signalling in 293T cells in a dose-dependent manner, demonstrated by the TOPFLASH/TK-*Renilla*

dual luciferase reporter assay (Fig. 4c and Extended Data Fig. 9a), as well as by phosphorylation levels of LRP6 (a FZD co-receptor) and DVL2 (a downstream Wnt signalling component) (Extended Data Fig. 9b)²⁹. TcdB₁₁₁₄₋₁₈₃₅ did not glucosylate small GTPases in colonic organoids (Extended Data Fig. 9c), yet it inhibited organoid growth and induced death (Fig. 4d and Extended Data Fig. 9d, e). The death of colonic organoids was rescued by CHIR99021, demonstrating that the effect of TcdB₁₁₁₄₋₁₈₃₅ is due to blockage of Wnt signalling. These data raised the intriguing possibility that binding of TcdB to FZDs may directly contribute to disruption of the colon epithelium by inhibiting Wnt signalling.

FZDs are TcdB receptors in colonic epithelium

Finally, we examined the colonic epithelium *in vivo*. Immunohistochemistry (IHC) analysis showed that FZD2 and FZD7 are expressed in mouse and human colonic epithelium (Fig. 5a, b and Extended Data Fig. 10a–f). In contrast, CSPG4 is predominantly expressed in the multi-nucleated intestinal sub-epithelial myofibroblasts (ISEMFs) and is not detectable in the colonic epithelium (Fig. 5c and Extended Data Fig. 10c), which is consistent with a previous report¹⁷.

As TcdB is released into the lumen of the colon during *C. difficile* infection, we developed a model in which we injected TcdB directly into the lumen of ligated colon segments in mice (Fig. 5d), which resulted in binding and entry of TcdB into the colonic epithelium (Fig. 5e). Co-injection of FZD2-CRD largely abolished binding of TcdB (Fig. 5e), suggesting that FZDs are the dominant receptors in the colonic epithelium.

To verify further the role of FZDs *in vivo*, we turned to FZD-knockout mouse models. *FZD2/7* double-knockout mice are embryonic lethal, and *FZD2*^{-/-} mice also displayed developmental defects⁴⁰. *FZD7*^{-/-} mice appear to develop normally and exhibit no overt intestinal defects under basal conditions^{39,40}. Thus, we chose *FZD7*^{-/-} mice to assess whether a loss of a major colonic FZD member may reduce TcdB toxicity *in vivo*. To focus the analysis on the colonic epithelium and avoid the potential effects of TcdB entry into CSPG4-expressing ISEMFs, we used TcdB₁₋₁₈₃₀ and injected the toxin into the lumen of ligated colon segments of live mice. After an 8-h incubation period, fluid accumulation was observed in wild-type mice, but was significantly reduced in *FZD7*^{-/-} mice (Fig. 5f). Histological scoring revealed extensive disruption of the epithelium, inflammatory cell infiltration and oedema in wild-type mice, but much less in *FZD7*^{-/-} mice (Fig. 5g and Extended Data Fig. 10g). To assess epithelial integrity further, we performed immunofluorescent analysis on colonic sections for the cell–cell junction markers claudin-3 and ZO-1. Both markers were extensively disrupted in wild-type mice after exposure to TcdB₁₋₁₈₃₀ but remained largely intact in *FZD7*^{-/-} mice (Fig. 5h and Extended Data Fig. 10h). Together, these data demonstrate that FZDs are physiologically relevant receptors for TcdB in the colonic epithelium *in vivo*.

Discussion

Our findings support a previously proposed two-receptor model for TcdB²⁵, but with a notable amendment: FZDs and CSPG4 may act as receptors in different cell types. CSPG4 is expressed in the ISEMFs, which are involved in diverse processes from wound healing to inflammation⁴¹. Although the role of ISEMFs in *C. difficile* infection remains to be established, it is conceivable that targeting these cells by TcdB could contribute to disease progression after FZD-mediated disruption of the colonic epithelium.

Our unbiased genome-wide screens revealed multiple host factors involved in all major steps of toxin actions, from receptors (FZDs and CSPG4) to acidification in endosomes (vacuolar-type H⁺-ATPase)^{42,43}, to enzymatic activity in the cytosol (UGP2) (Extended Data Fig. 10i). Many other top-ranking hits remain to be validated, such as FBXO11 and enzymes involved in phospholipid metabolism/signalling,

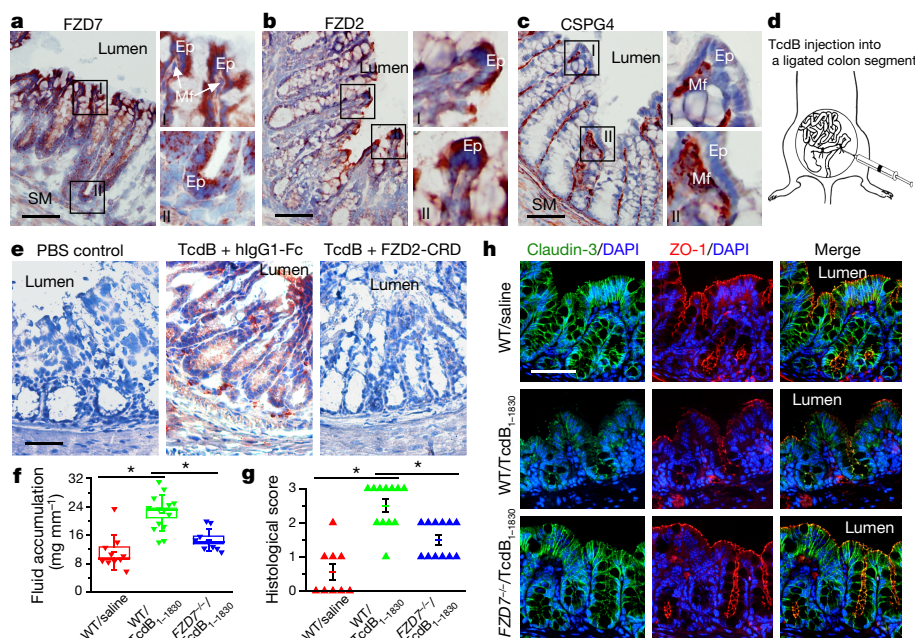


Figure 5 | FZDs are TcdB receptors in the colonic epithelium.

a–c, Mouse colon cryosections were subjected to IHC analysis to detect FZD7 (**a**), FZD2 (**b**) and CSPG4 (**c**). Blue indicates cell nuclei, red indicates target proteins. Ep, epithelial cells; Mf, sub-epithelial myofibroblasts; SM, smooth muscles. **d**, A schematic illustration of the colon loop ligation assay. **e**, Co-injection of FZD2-CRD with TcdB into the ligated colonic segments prevented TcdB binding to the colonic epithelium, analysed by IHC. Red indicates TcdB and blue indicates cell nuclei. **f–h**, TcdB_{1–1830} was injected into the ligated colonic segments and incubated for 8 h in wild-type (WT)

and FZD7^{−/−} mice. The colon segments were then excised and analysed for fluid accumulation (**f**), histological scoring by haematoxylin and eosin (H&E) staining (**g**; the representative images are in Extended Data Fig. 10g) and for the cell–cell junction markers claudin-3 and ZO-1 by immunofluorescent staining (**h**; blue indicates cell nuclei). DAPI, 4',6-diamidino-2-phenylindole. Each data point represents one mouse (**f**, **g**). Boxes represent mean ± standard error of the mean (s.e.m.), and the bars represent s.d. (**P* < 0.005, one-way ANOVA). Scale bars, 50 μm. Representative images are from one of three independent experiments.

including phosphatidylinositol 5-phosphate 4-kinase (PIP4K2B), phosphatidylinositol 4-kinase (PI4KB) and phospholipase C (PLCG1) (Extended Data Fig. 10i).

Our screen identified many key players in Wnt signalling pathways, including APC, GSK-3β, WNT5A and LRP6 (Extended Data Fig. 10i). It has been suggested that TcdA attenuates Wnt signalling in cells, although the effects appear to be indirect, largely due to deactivation of Rho GTPase by TcdA⁴⁴. Wnt signalling is particularly important for maintaining colonic stem cells^{30,45}, which continuously give rise to new colonic epithelial cells. The health of these stem cells is critical for self-renewal and repair of the colonic epithelium, which has an extraordinarily fast turnover rate⁴⁵. Our findings suggest that colonic stem cells are a major target of TcdB. The potential role of Wnt signalling inhibition in the pathogenesis of *C. difficile* infection, and the therapeutic potential of modulating Wnt signalling downstream of FZDs warrant further study. Finally, dysregulation of Wnt signalling pathways is associated with many cancers, particularly colorectal cancers^{30,46}. The receptor-binding domain of TcdB, or its homologues, may serve as valuable tools and potential therapeutics for targeting Wnt signalling pathways.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 April; accepted 26 August 2016.

Published online 28 September 2016.

1. Lyster, D. M., Krivan, H. C. & Wilkins, T. D. *Clostridium difficile*: its disease and toxins. *Clin. Microbiol. Rev.* **1**, 1–18 (1988).
2. Rupnik, M., Wilcox, M. H. & Gerding, D. N. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature Rev. Microbiol.* **7**, 526–536 (2009).
3. Heinlen, L. & Ballard, J. D. *Clostridium difficile* infection. *Am. J. Med. Sci.* **340**, 247–252 (2010).
4. Voth, D. E. & Ballard, J. D. *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin. Microbiol. Rev.* **18**, 247–263 (2005).

5. Hunt, J. J. & Ballard, J. D. Variations in virulence and molecular biology among emerging strains of *Clostridium difficile*. *Microbiol. Mol. Biol. Rev.* **77**, 567–581 (2013).
6. Lessa, F. C. *et al.* Burden of *Clostridium difficile* infection in the United States. *N. Engl. J. Med.* **372**, 825–834 (2015).
7. Jank, T. & Aktories, K. Structure and mode of action of clostridial glucosylating toxins: the ABCD model. *Trends Microbiol.* **16**, 222–229 (2008).
8. Sun, X., Savidge, T. & Feng, H. The enterotoxicity of *Clostridium difficile* toxins. *Toxins (Basel)* **2**, 1848–1880 (2010).
9. Pruitt, R. N. & Lacy, D. B. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front. Cell. Infect. Microbiol.* **2**, 28 (2012).
10. Just, I. *et al.* Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375**, 500–503 (1995).
11. Drudy, D., Fanning, S. & Kyne, L. Toxin A-negative, toxin B-positive *Clostridium difficile*. *Int. J. Infect. Dis.* **11**, 5–10 (2007).
12. Lyras, D. *et al.* Toxin B is essential for virulence of *Clostridium difficile*. *Nature* **458**, 1176–1179 (2009).
13. Kuehne, S. A. *et al.* The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* **467**, 711–713 (2010).
14. Carter, G. P. *et al.* Defining the roles of TcdA and TcdB in localized gastrointestinal disease, systemic organ damage, and the host response during *Clostridium difficile* infections. *MBio* **6**, e00551 (2015).
15. Hamm, E. E., Voth, D. E. & Ballard, J. D. Identification of *Clostridium difficile* toxin B cardiotoxicity using a zebrafish embryo model of intoxication. *Proc. Natl Acad. Sci. USA* **103**, 14176–14181 (2006).
16. Yuan, P. *et al.* Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. *Cell Res.* **25**, 157–168 (2015).
17. Terada, N. *et al.* Immunohistochemical study of NG2 chondroitin sulfate proteoglycan expression in the small and large intestines. *Histochem. Cell Biol.* **126**, 483–490 (2006).
18. LaFrance, M. E. *et al.* Identification of an epithelial cell receptor responsible for *Clostridium difficile* TcdB-induced cytotoxicity. *Proc. Natl Acad. Sci. USA* **112**, 7073–7078 (2015).
19. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
20. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
21. Greco, A. *et al.* Carbohydrate recognition by *Clostridium difficile* toxin A. *Nature Struct. Mol. Biol.* **13**, 460–461 (2006).
22. Barroso, L. A., Moncrief, J. S., Lyster, D. M. & Wilkins, T. D. Mutagenesis of the *Clostridium difficile* toxin B gene and effect on cytotoxic activity. *Microb. Pathog.* **16**, 297–303 (1994).

23. Genisyuerek, S. *et al.* Structural determinants for membrane insertion, pore formation and translocation of *Clostridium difficile* toxin B. *Mol. Microbiol.* **79**, 1643–1654 (2011).
24. Olling, A. *et al.* The repetitive oligopeptide sequences modulate cytopathic potency but are not crucial for cellular uptake of *Clostridium difficile* toxin A. *PLoS ONE* **6**, e17623 (2011).
25. Schorch, B. *et al.* LRP1 is a receptor for *Clostridium perfringens* TpeL toxin indicating a two-receptor model of clostridial glycosylating toxins. *Proc. Natl Acad. Sci. USA* **111**, 6431–6436 (2014).
26. Ryder, A. B. *et al.* Assessment of *Clostridium difficile* infections by quantitative detection of *tcdB* toxin by use of a real-time cell analysis system. *J. Clin. Microbiol.* **48**, 4129–4134 (2010).
27. Flores-Díaz, M. *et al.* Cellular UDP-glucose deficiency caused by a single point mutation in the UDP-glucose pyrophosphorylase gene. *J. Biol. Chem.* **272**, 23784–23791 (1997).
28. Chaves-Olarte, E. *et al.* UDP-glucose deficiency in a mutant cell line protects against glucosyltransferase toxins from *Clostridium difficile* and *Clostridium sordellii*. *J. Biol. Chem.* **271**, 6925–6932 (1996).
29. MacDonald, B. T. & He, X. Frizzled and LRP5/6 receptors for Wnt/ β -catenin signaling. *Cold Spring Harb. Perspect. Biol.* **4**, a007880 (2012).
30. Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* **19**, 877–890 (2005).
31. Jonikas, M. C. *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
32. Christianson, J. C. *et al.* Defining human ERAD networks through an integrative mapping strategy. *Nature Cell Biol.* **14**, 93–105 (2011).
33. Orth, P. *et al.* Mechanism of action and epitopes of *Clostridium difficile* toxin B-neutralizing antibody bezlotoxumab revealed by X-ray crystallography. *J. Biol. Chem.* **289**, 18008–18021 (2014).
34. Sagara, N., Toda, G., Hirai, M., Terada, M. & Katoh, M. Molecular cloning, differential expression, and chromosomal localization of human *Frizzled-1*, *Frizzled-2*, and *Frizzled-7*. *Biochem. Biophys. Res. Commun.* **252**, 117–122 (1998).
35. Richard, M., Boulin, T., Robert, V. J., Richmond, J. E. & Bessereau, J. L. Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. *Proc. Natl Acad. Sci. USA* **110**, E1055–E1063 (2013).
36. Satoh, T., Ohba, A., Liu, Z., Inagaki, T. & Satoh, A. K. dPob/EMC is essential for biosynthesis of rhodopsin and other multi-pass membrane proteins in *Drosophila* photoreceptors. *eLife* **4**, e06306 (2015).
37. Ueno, K. *et al.* Frizzled-7 as a potential therapeutic target in colorectal cancer. *Neoplasia* **10**, 697–705 (2008).
38. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
39. Flanagan, D. J. *et al.* Frizzled7 functions as a Wnt receptor in intestinal epithelial Lgr5⁺ stem cells. *Stem Cell Reports* **4**, 759–767 (2015).
40. Yu, H., Ye, X., Guo, N. & Nathans, J. Frizzled 2 and Frizzled 7 function redundantly in convergent extension and closure of the ventricular septum and palate: evidence for a network of interacting genes. *Development* **139**, 4383–4394 (2012).
41. Powell, D. W. *et al.* Myofibroblasts. II. Intestinal subepithelial myofibroblasts. *Am. J. Physiol.* **277**, C183–C201 (1999).
42. Qa'Dan, M., Spyres, L. M. & Ballard, J. D. pH-induced conformational changes in *Clostridium difficile* toxin B. *Infect. Immun.* **68**, 2470–2474 (2000).
43. Pfeifer, G. *et al.* Cellular uptake of *Clostridium difficile* toxin B. Translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells. *J. Biol. Chem.* **278**, 44535–44541 (2003).
44. Bezerra Lima, B. *et al.* *Clostridium difficile* toxin A attenuates Wnt/ β -catenin signaling in intestinal epithelial cells. *Infect. Immun.* **82**, 2680–2687 (2014).
45. Crosnier, C., Stamatakis, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature Rev. Genet.* **7**, 349–359 (2006).
46. Gujral, T. S. *et al.* A noncanonical Frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis. *Cell* **159**, 844–856 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Dong laboratory, L. Peng, X. Zhong, Q. Ma and M. Waldor for discussions; B. Ding and Y. Jing for their assistance in data analysis; N. Renzette and T. Kowalik for their advice and access to the NGS sequencer; H. Tatge for assistance on constructing toxin-expression plasmids; J. Nathans for providing FZD7 and FZD8-CRD-Myc-GPI constructs. This study was supported by National Institutes of Health (NIH) grants R01NS080833 (M.D.), R01AI091786 (A.L.B.), R01AT006732 (J.H.), R01DK084056 (D.T.B.), R01CA095287 (W.B.S.), K99DK100539 (J.M.), R01GM057603, R01GM074241 and R01AR060359 (X.H.). We also acknowledge support from the Bill and Melinda Gates Foundation (P.M. and A.L.B.), the Timothy Murphy Fund (D.T.B.), the Harvard Digestive Diseases Center (NIH P30DK034854, X.H., D.T.B. and M.D.), and the Boston Children's Hospital Intellectual and Developmental Disabilities Research Center (NIH P30HD18655, X.H., D.T.B. and M.D.). X.H. is an American Cancer Society Research Professor. M.D. and A.L.B. both hold the Investigator in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund.

Author Contributions L.T., A.L.B. and M.D. conceived the project. L.T. designed and conducted the majority of the experiments. J.Z. and L.T. developed the colon loop ligation model and conducted *in vivo* experiments. P.M. and A.L.B. prepared the CRISPR library and cells for screening and contributed to the screen design and data analysis. A.T. and D.T.B. prepared colonic organoids, performed immunofluorescent experiments, and analysed data. J.M. prepared adenoviruses and assisted with knockdown experiments. X.Z. and X.H. assisted with the Wnt signalling inhibition experiments and data analysis. W.B.S. provided key reagents/advice on CSPG4. R.G., X.W. and J.G.H. provided key advice/reagents on TcdB and TcdA. L.T. and M.D. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D. (min.dong@childrens.harvard.edu).

Reviewer Information Nature thanks J. Ballard, N. Fairweather and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cell lines, antibodies and constructs. HeLa (H1, #CRL-1958), CHO (K1, #CCL-61), HT-29 (#HTB-38), Caco-2 (#HTB-37) and 293T (#CRL-3216) cells were originally obtained from ATCC. They tested negative for mycoplasma contamination, but have not been authenticated. The following mouse monoclonal antibodies were purchased from the indicated vendors: RAC1 (23A8, Abcam), non-glucosylated RAC1 (Clone 102, BD Biosciences), 1D4 tag (MA1-722, ThermoFisher Scientific), HA tag (16B12, Covance), β -actin (AC-15, Sigma), ZO-1 (339100, Life technology). Rabbit monoclonal IgG against human CSPG4 (ab139406) and rabbit polyclonal antibodies against FZD1 (ab150553), FZD2 (ab150477), FZD7 (ab51049), PVRL3 (ab63931) and claudin-3 (ab15102) were all purchased from Abcam. Rabbit monoclonal antibodies against DVL2 (30D2) and LRP6 (C5C7), and a rabbit polyclonal antibody against phosphorylated LRP6 (Ser1490) were all purchased from Cell Signaling. Chicken polyclonal IgY (#754A) against TcdB was purchased from List Biological Labs. Antibody validation is available on the manufacturers' websites. A rabbit polyclonal antibody against rodent CSPG4 and a construct expressing full-length rat CSPG4 (in pcDNA vector) were both generated in W. Stallcup's laboratory. 1D4-tagged full-length FZD1-10 constructs (in pRK5 vector) were originally generated in J. Nathans' laboratory (Baltimore, MD) and were obtained from Addgene. FZD7 and FZD8-CRD-Myc-GPI constructs were generously provided by J. Nathans and have been described previously⁴⁷. Constructs expressing full-length human IL1RAPL2 and full-length PVRL3 were purchased from Vigene Biosciences. A construct expressing full-length mouse Syt II was described previously⁴⁸.

TcdB and other recombinant proteins. Recombinant TcdB (from *C. difficile* strain VPI 10463) and TcdA were expressed in *Bacillus megaterium* as previously described⁴⁹ and purified as His6-tagged proteins. TcdB₁₋₁₈₃₀ was cloned into pHIS1522 vector (MoBiTec) and expressed in *B. megaterium*. TcdB₁₈₃₁₋₂₃₆₆, TcdB₁₅₀₁₋₂₃₆₆ and TcdB₁₁₁₄₋₁₈₃₅ were cloned into pGEX-6P-1 or pET28a vectors and purified as GST-tagged or His6-tagged proteins in *Escherichia coli*. Rat CSPG4-EC (pool (P1) and P2) was expressed in HEK293 cells, purified from medium with DEAE-Sepharose columns, and eluted with a gradient buffer (NaCl from 0.2 to 0.8 M, 50 mM Tris-Cl, pH 8.6) as previously described⁵⁰. Recombinant human proteins were purchased from ACRO Biosystems (IgG1 Fc and FZD2-CRD-Fc), R&D Systems (FZD1-CRD-Fc, FZD5-CRD-Fc and FZD7-CRD-Fc), Sino Biologicals (PVRL3-EC), and StemRD (WNT3A).

Generating stable HeLa-Cas9 cells and lentivirus sgRNA libraries. The human codon-optimized sequence of *S. pyogenes* Cas9 was subcloned from plasmid lenti-Cas9-Blast (Addgene #52962) into the pQCXIH retroviral vector (Clontech), which was used to generate retroviruses to transduce HeLa cells. Mixed stable cells were selected in the presence of hygromycin B (200 μ g/ml, Life Technologies). Lentivirus sgRNA libraries were generated following published protocols using the human GeCKO v.2 sgRNA library (Addgene #1000000049)¹⁹. The GeCKO v.2 library is composed of two half-libraries (library A and library B). Each half-library contains three unique sgRNA per gene and was independently screened with toxins. Cells were transduced with lentivirus-packaged sgRNA library at a MOI of 0.2.

Screening CRISPR libraries with TcdB and TcdB₁₋₁₈₃₀. For each CRISPR half-library of cells, 4×10^7 cells were plated onto two 15-cm culture dishes to ensure sufficient coverage of sgRNAs, with each sgRNA on average being represented about 650 times (that is, there are on average 650 cells transduced with the same sgRNA). This over-representation rate was calculated from titration plates that were set up in parallel with the library. These cells were exposed to either TcdB or TcdB₁₋₁₈₃₀ for 48 h. Cells were then washed three times with PBS to remove loosely attached round-shaped cells. The remaining cells were re-seeded and cultured with normal medium without toxins until ~70% confluence. Cells were then subjected to the next round of screening with increased concentrations of toxins. Four rounds of screenings were carried out with TcdB (0.05, 0.1, 0.2 and 0.5 pM) and TcdB₁₋₁₈₃₀ (5, 10, 20 and 50 pM). The remaining cells were harvested and their genomic DNA extracted using the Blood and Cell Culture DNA mini kit (Qiagen). DNA fragments containing the sgRNA sequences were amplified by PCR using primers lentiGP-1_F (AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCG) and lentiGP-3_R (ATGAATACTGCCATTTGTCTCAAGATCTAGTTACGC). NGS (Illumina MiSeq) was performed by a commercial vendor (Genewiz).

Generating HeLa knockout cell lines. The following sgRNA sequences were cloned into LentiGuide-Puro vectors (Addgene) to target the indicated genes: CCGGAGACACGGAGCAGTGG (CSPG4), CGCGTGTGGGACATCGCCT (EMC4), ACCTTATACCACACATC (IL1RAPL2), TGCGAGCATTCC CGCGCCA (FZD2), AGCGCATGACCACTACTG (SGMS1), ACAGGCA GAAAACGGCTCCT (UGP2), GTGTAATGACAAGTTCGCCG (FZD1), and GAGAACGGTAAAGAGCGTCG (FZD7). HeLa-Cas9 cells were transduced with lentiviruses that express these sgRNAs. Mixed populations of stable cells were selected with puromycin (2.5 μ g/ml) and hygromycin B (200 μ g/ml).

FZD1/2/7^{-/-} cells were created by sequentially transducing FZD1 and FZD7 sgRNA lentiviruses into FZD2^{-/-} cells and further selected in the presence of 100 pM TcdB₁₋₁₈₃₀. The mutagenesis rate in these mixed stable cells was determined by NGS (Supplementary Table 1).

Cytopathic assay. The cytopathic effect (cell rounding) of TcdA and TcdB was analysed using standard cell-rounding assay as previously described¹. Briefly, cells were exposed to a gradient of TcdB and TcdB₁₋₁₈₃₀ for 24 h. Phase-contrast images of cells were taken (Olympus IX51, $\times 10$ –20 objectives). The numbers of round-shaped and normal shaped cells were counted manually. The percentage of round-shaped cells was plotted and fitted using the Origin software.

Blocking TcdB entry into cells with CSPG4-EC and FZD2-CRD-Fc. Recombinant proteins used for cell protection assays were pre-filtered (0.22 μ m, Millipore). Toxins were pre-incubated with FZD2-CRD-Fc and/or CSPG4-EC (P1) for 30 min on ice with a toxin/protein ratio of 1:400 (except when specifically noted in the figure legend). The mixtures were added into cell culture medium and cells were analysed by the cytopathic assay.

Transfection, TcdB binding to cells, and immunoblot analysis. Transient transfection of HeLa cells was carried out using PolyJet (SigmaGen). Binding of TcdB to cells was analysed by exposing cells to TcdB or truncated TcdB fragments for 10 min at room temperature. Cells were washed three times with PBS and then either fixed for immunostaining analysis or harvested with RIPA buffer (50 mM Tris, 1% NP40, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, plus a protease inhibitor cocktail (Sigma-Aldrich)). Cell lysates were centrifuged and supernatants were subjected to SDS-PAGE and immunoblot analysis using the enhanced chemiluminescence method (Pierce). The full blot images are shown in Supplementary Fig. 1.

Pulldown assays. Pulldown assays were carried out using glutathione Sepharose 4B as previously described⁴⁸. Briefly, 5 μ g of GST-tagged TcdB₁₈₃₁₋₂₃₆₆ and TcdB₁₅₀₁₋₂₃₆₆ were immobilized on glutathione beads and incubated with FZD2-CRD-Fc (10 nM) for 1 h at 4°C. Beads were then washed, pelleted, boiled in SDS sample buffer, and subjected to immunoblot analysis.

BLI assay. The binding affinities between TcdB and FZD-CRDs were measured by BLI assay using the Blitz system (ForteBio). Briefly, the CRDs-Fc of FZD1, 2, 5, 7 or human IgG1 Fc (20 μ g/ml) were immobilized onto capture biosensors (Dip and Read Anti-IgG-Fc, ForteBio) and balanced with PBS. The biosensors were then exposed to TcdB or TcdB₁₋₁₈₃₀, followed by washing with PBS. Binding affinities (K_d) were calculated using the Blitz system software (ForteBio).

Wnt signalling assay. The TOPFLASH/TK-*Renilla* dual luciferase reporter assay was used to detect Wnt signalling activities as previously described⁵¹. Briefly, 293T cells in 24-well plates were co-transfected with TOPFLASH (50 ng/well), TK-*Renilla* (internal control, 10 ng/well), and pcDNA3 (200 ng/well). After 24 h, cells were exposed to WNT3A (50 ng/ml) and TcdB₁₁₁₄₋₁₈₃₅ (1:8, 1:40, and 1:200 to WNT3A) in culture medium for 6 h. Cell lysates were harvested and subjected to either firefly/*Renilla* dual luciferase assay or immunoblot analysis for detecting phosphorylated DVL2 and LRP6. Wnt signalling activates expression of TOPFLASH luciferase reporter (firefly luciferase). Co-transfected *Renilla* luciferase serves as an internal control.

Microtitre plate-based binding assay. Binding assays were performed on 96-well plates (EIA/RIA plate, Corning Costar) as described previously⁵⁰. Briefly, microtitre plates were coated with 10 μ g/ml rat CSPG4-EC proteins in coating buffer (0.1 M NaHCO₃, pH 8.3) at 4°C overnight, and then blocked with 1% bovine serum albumin in PBS for 1 h. Plates were then incubated with the indicated proteins for 1 h in PBS. Wells were washed three times with PBS plus 0.05% Tween-20 at room temperature. One-step Turbo TMB (ThermoFisher Scientific) was used as the substrate, and absorbance at 450 nm was measured with a microplate reader.

Organoid culture, knockdown, and TcdB challenge assay. Crypt isolation from wild-type or FZD7^{-/-} mouse colon was carried out as previously described, and organoids were expanded as spheroid cultures using conditioned medium⁵². Except for wild-type organoids used for Wnt signalling inhibition assay, CHIR99021 (3 μ M) was also added to the medium³⁹. Five days after passaging, organoids were re-suspended with Cell Recovery Solution (ThermoFisher Scientific) and mechanically fragmented. Fragments were transduced with adenoviruses expressing shRNA for FZD1, FZD2, or a control shRNA sequence using medium supplemented with Nicotinamide (10 mM, Sigma), Polybrene (8 μ g/ml, Sigma), and Y-27632 (10 μ M, Sigma), washed, and plated in growth factor-reduced Matrigel (Corning)⁵³. Three days following viral transduction, organoids were challenged with TcdB by adding the toxin into the medium. Viability of organoids was quantified after 72 h.

Wnt signalling inhibition in wild-type colon organoids. TcdB₁₁₁₄₋₁₈₃₅ was added into the culture medium of wild-type colon organoids. For rescue experiments, 5 μ M CHIR99021 was also added to the medium. The medium was changed every 48 h with the constant presence of TcdB₁₁₁₄₋₁₈₃₅ and/or CHIR99021. Viability of cells was analysed after 6 days.

Generating FZD1 and FZD2-knockdown adenovirus. All shRNAs were purchased from Sigma (MISSION shRNA library). The knockdown efficiency was validated as described in Extended Data Fig. 8c, d. shRNA sequences showing the highest efficiency were selected to generate adenoviruses. Adenoviruses expressing a control shRNA (5'-CTGGACTTCCAGAAGAACA-3'), shRNAs against mouse FZD1 (shRNA#2: 5'-TGGTGTGCAACGACAAGTTTG-3'), or FZD2 (shRNA#5: 5'-CGCTTCTCAGAGGACGGTTAT-3') were constructed using the Block-it U6 adenoviral RNAi system (Life Technologies), followed by viral packaging and multiple rounds of amplification in 293A cells (Life Technologies).

Viability assay for colonic organoids. The viability of colonic organoids was assessed using the MTT assay as previously described⁵⁴. Briefly, the MTT solution was added to the organoid culture (500 µg/ml). After incubation at 37 °C for 2 h, the medium was discarded. For each well (containing 20 µl of Matrigel, in a 48-well plate), 60 µl of 2% SDS solution was added to solubilize the Matrigel (1 h, 37 °C), followed by the addition of 300 µl of dimethylsulfoxide (DMSO) to solubilize reduced MTT (2 h, 37 °C). The absorbance at 562 nm was measured on a microplate reader. Twenty microlitres of Matrigel without organoids was used as blank control. Normal organoids without exposure to toxins were considered as 100% viable.

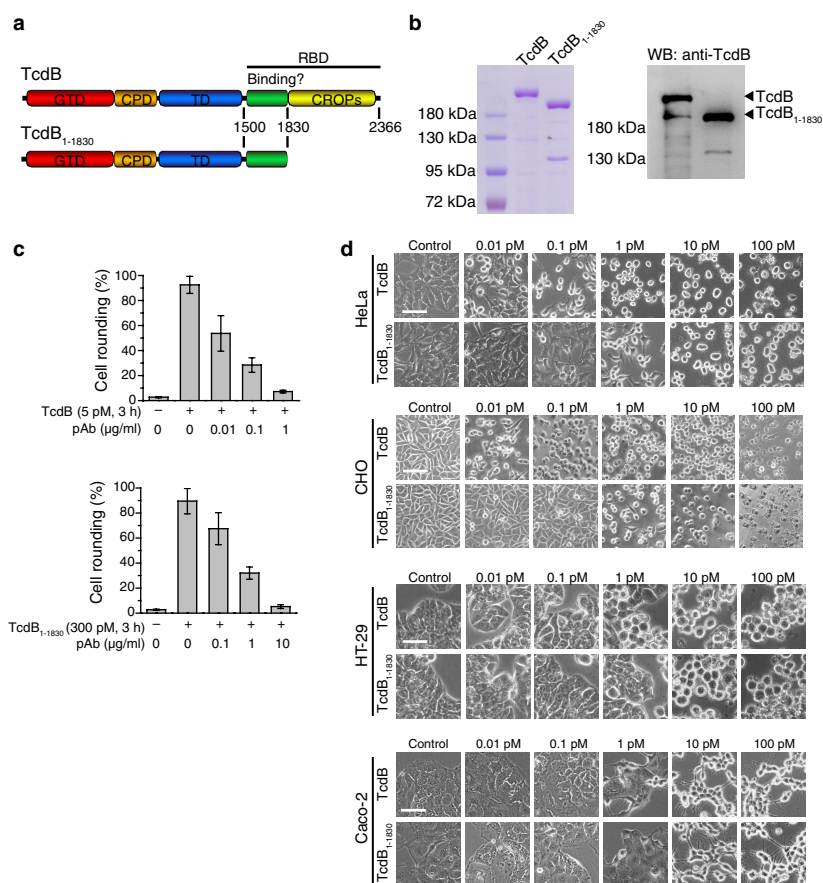
IHC, immunofluorescence and histology analysis. Colons from adult mice (C57BL/6 strain (purchased from The Jackson Laboratory, #000664), 10–12 weeks old, both male and female mice were used and randomly distributed into experimental groups) were dissected out and subjected to cryosectioning into sections 8–10 µm thick. Colonic sections were fixed in cold acetone for 5 min and then washed three times with PBS. The colonic sections were then blocked with 5% goat serum in PBS for 30 min at room temperature and incubated with primary antibodies overnight (anti-TcdB: 1:600; anti-FZDs: 1:250; rabbit anti-CSPG4: 1:250), followed with biotinylated goat anti-chicken or rabbit IgG secondary antibodies (1:200, Vector Laboratory) for 1 h at room temperature. The sections were then incubated with horseradish peroxidase (HRP)-conjugated streptavidin (1:500, DAKO) for 30 min. Immunoreactivity was visualized as red colour with 3-amino-9-thyl carbazole (DAKO). Cell nuclei were labelled blue with Gill's haematoxylin (1:3.5, Sigma). Frozen human colon tissue slides were purchased from BioChain Institute and subjected to IHC analysis. Immunofluorescence analysis of claudin-3 and ZO-1 was carried out using mouse colon tissues fixed in 10% formalin and embedded in paraffin (anti-claudin-3: 1:100; anti-ZO-1: 1:100). Confocal images were captured with the Ultraview Vox Spinning Disk Confocal System. Histology analysis was carried out with H&E staining of paraffin-embedded sections. Stained sections were coded and scored by observers blinded to experimental groups, based on disruption of the colonic epithelium, inflammatory cell infiltration and oedema, on a scale of 0 to 3 (mild to severe). No statistical methods were used to predetermine sample size.

Competition assays in colonic tissues with recombinant proteins. All procedures were conducted in accordance with the guidelines approved by the Boston

Children's Hospital Institutional Animal Care and Use Committee (IACUC) (#3028). TcdB (40 nM) was pre-incubated with either human IgG1-Fc or FZD2-Fc (2.4 µM) for 30 min on ice. To generate the *ex vivo* colon segments, mice (C57BL/6, 6–8 weeks, both male and female mice were used, repeated three times, each time four mice per group, the experiments were not randomized or blinded) were euthanized and the colon exposed via laparotomy. A segment in the ascending colon (~2 cm long) was sealed by tying both ends with silk ligatures. The toxin samples (40 µl) were injected through an intravenous catheter into the sealed colon segment. The injection site was then sealed with a haemostat. The colon was covered with PBS-soaked gauze for 2 h, then excised and its lumen flushed with PBS three times, and subjected to IHC analysis.

Colon loop ligation assay. All procedures were conducted in accordance with the guidelines approved by the Boston Children's Hospital IACUC (#3028). Wild-type or *FZD7^{-/-}* mice (The Jackson Laboratory, #012825, strain B6;129-Fzd7^{tm1.1Nat/J}), 6–8 weeks old, sample size indicated in Fig. 5f, g, both male and female mice were used, the experiments were not randomized or blinded) were anaesthetized following overnight fasting. A midline laparotomy was performed to locate the ascending colon and seal a ~2 cm loop with silk ligatures. Two micrograms of TcdB_{1–1830} in 80 µl of normal saline or 80 µl of normal saline were injected through an intravenous catheter into the sealed colon segment, followed by closing the wounds with stitches. Mice were allowed to recover. After 8 h, mice were euthanized and the ligated colon segments were excised, weighed, and measured. The colon segments were fixed, paraffin-embedded, sectioned, and subjected to either H&E staining for histological score analysis or immunofluorescent staining for claudin-3 and ZO-1.

47. Hsieh, J. C., Rattner, A., Smallwood, P. M. & Nathans, J. Biochemical characterization of Wnt-Frizzled interactions using a soluble, biologically active vertebrate Wnt protein. *Proc. Natl Acad. Sci. USA* **96**, 3546–3551 (1999).
48. Dong, M. *et al.* Synaptotagmins I and II mediate entry of botulinum neurotoxin B into cells. *J. Cell Biol.* **162**, 1293–1303 (2003).
49. Yang, G. *et al.* Expression of recombinant *Clostridium difficile* toxin A and B in *Bacillus megaterium*. *BMC Microbiol.* **8**, 192 (2008).
50. Tillet, E., Ruggiero, F., Nishiyama, A. & Stallcup, W. B. The membrane-spanning proteoglycan NG2 binds to collagens V and VI through the central nonglobular domain of its core protein. *J. Biol. Chem.* **272**, 10769–10776 (1997).
51. MacDonald, B. T., Yokota, C., Tamai, K., Zeng, X. & He, X. Wnt signal amplification via activity, cooperativity, and regulation of multiple intracellular PPPSP motifs in the Wnt co-receptor LRP6. *J. Biol. Chem.* **283**, 16115–16123 (2008).
52. Miyoshi, H. & Stappenbeck, T. S. *In vitro* expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nature Protocols* **8**, 2471–2482 (2013).
53. Wang, N. *et al.* Adenovirus-mediated efficient gene transfer into cultured three-dimensional organoids. *PLoS ONE* **9**, e93608 (2014).
54. Grabinger, T. *et al.* Ex vivo culture of intestinal crypt organoids as a model system for assessing cell death induction in intestinal epithelial cells and enteropathy. *Cell Death Dis.* **5**, e1228 (2014).

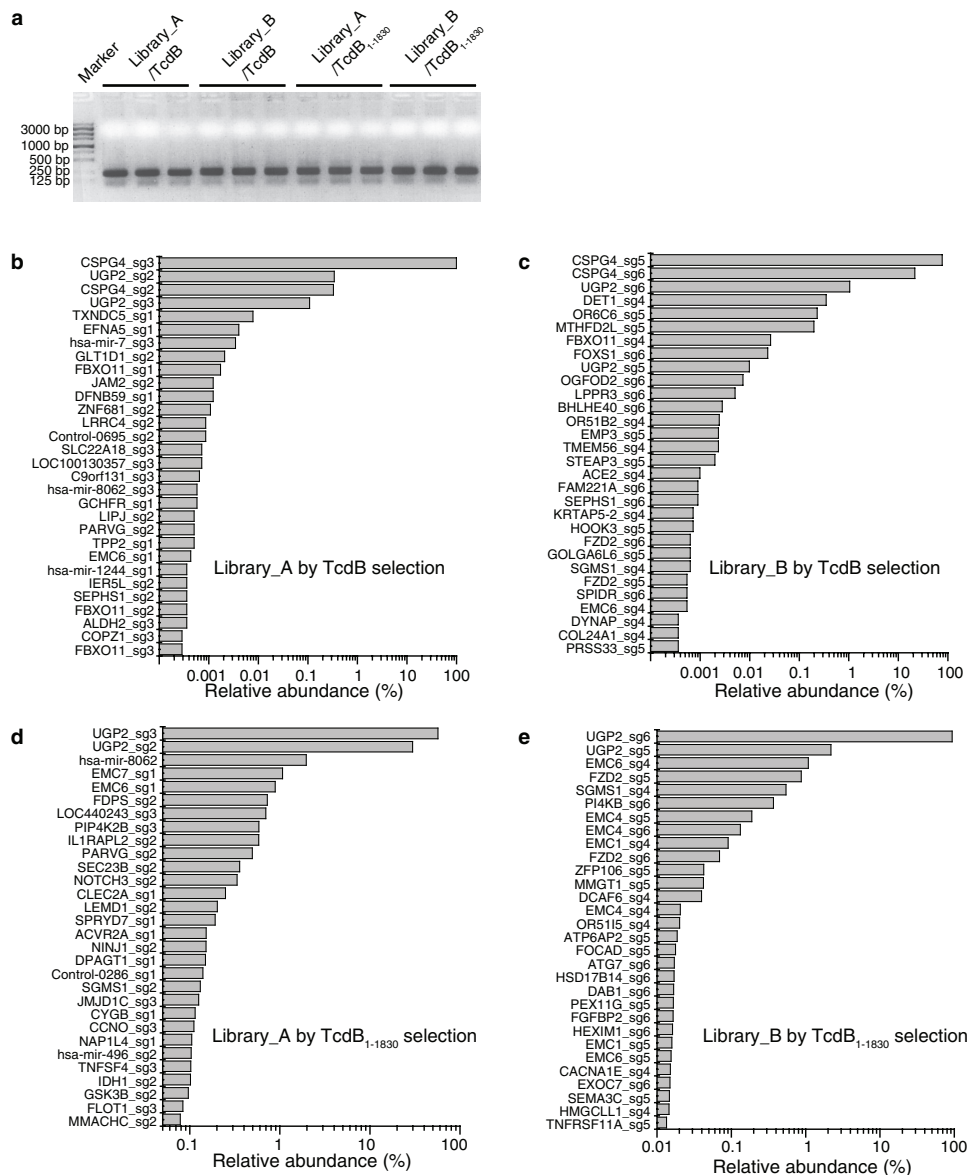


Extended Data Figure 1 | Recombinant TcdB and TcdB₁₋₁₈₃₀

a, Schematic drawings of TcdB and a truncated TcdB lacking the CROP region (TcdB₁₋₁₈₃₀). CPD, cysteine protease domain; GTD, glucosyltransferase domain; RBD, receptor-binding domain, including a putative receptor-binding region and the CROPs region; TD, translocation domain. **b**, Coomassie blue staining (left) and immunoblot (right; chicken polyclonal TcdB antibody) showing TcdB and TcdB₁₋₁₈₃₀ recombinantly expressed in *Bacillus megaterium*. We note that TcdB₁₋₁₈₃₀ contains a contaminating protein visible on Coomassie blue-stained gel. Mass spectrometry analysis confirmed that this band is not a fragment of TcdB.

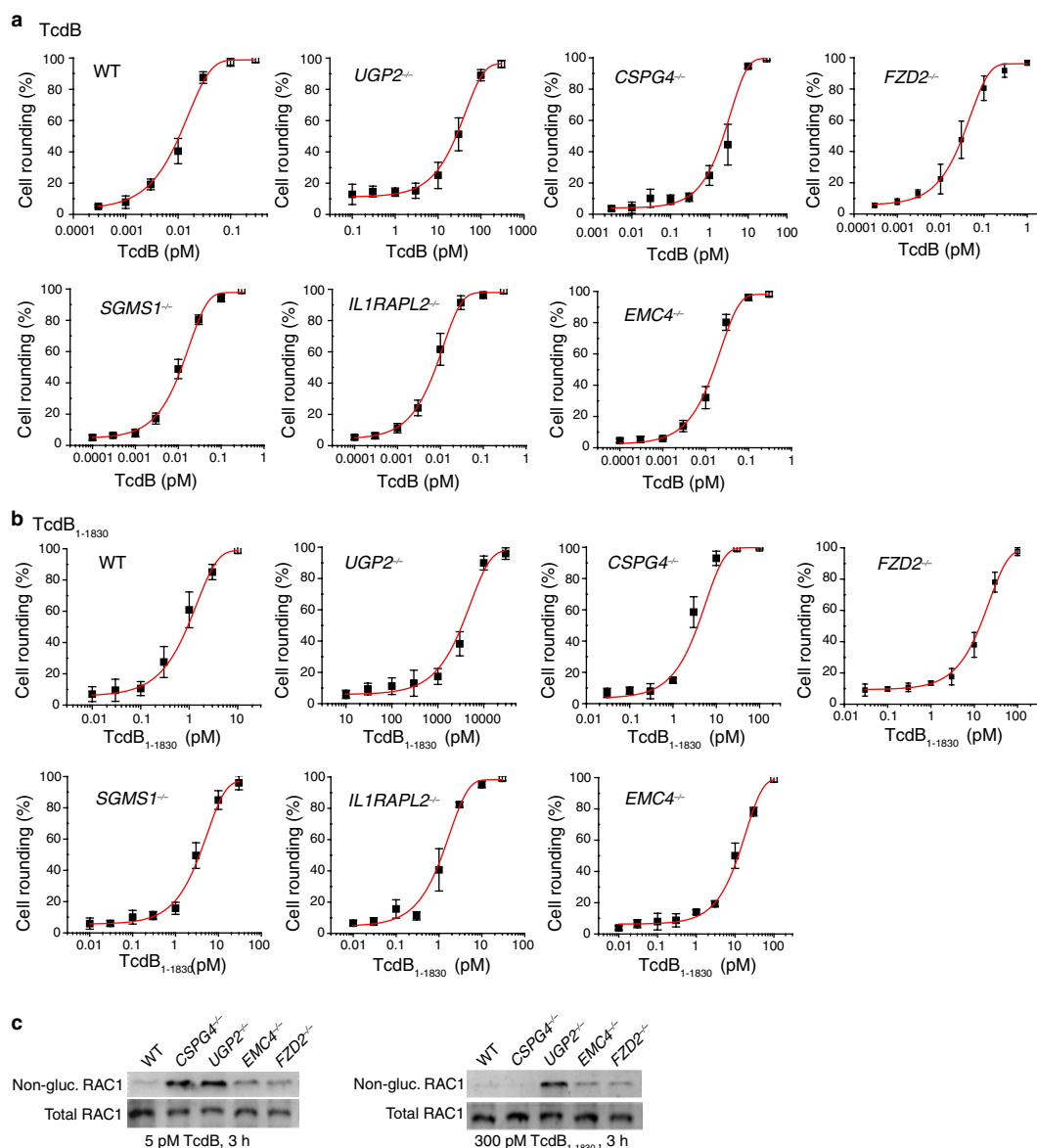
The top matching protein is the bacterial chaperone protein ClpB.

c, Cytopathic toxicity of recombinant TcdB and TcdB₁₋₁₈₃₀ on HeLa cells was neutralized by anti-TcdB polyclonal antibody (pAb), confirming that the toxicity is from TcdB and TcdB₁₋₁₈₃₀ (error bars indicate mean ± s.d., two independent experiments). **d**, HeLa, CHO, HT-29, and Caco-2 cells were exposed to TcdB or TcdB₁₋₁₈₃₀ as indicated for 24 h. TcdB₁₋₁₈₃₀ induced cell rounding at picomolar concentrations. Scale bars: 25 μm (HT-29) or 50 μm (HeLa, CHO and Caco-2). Representative images are from one of three independent experiments.



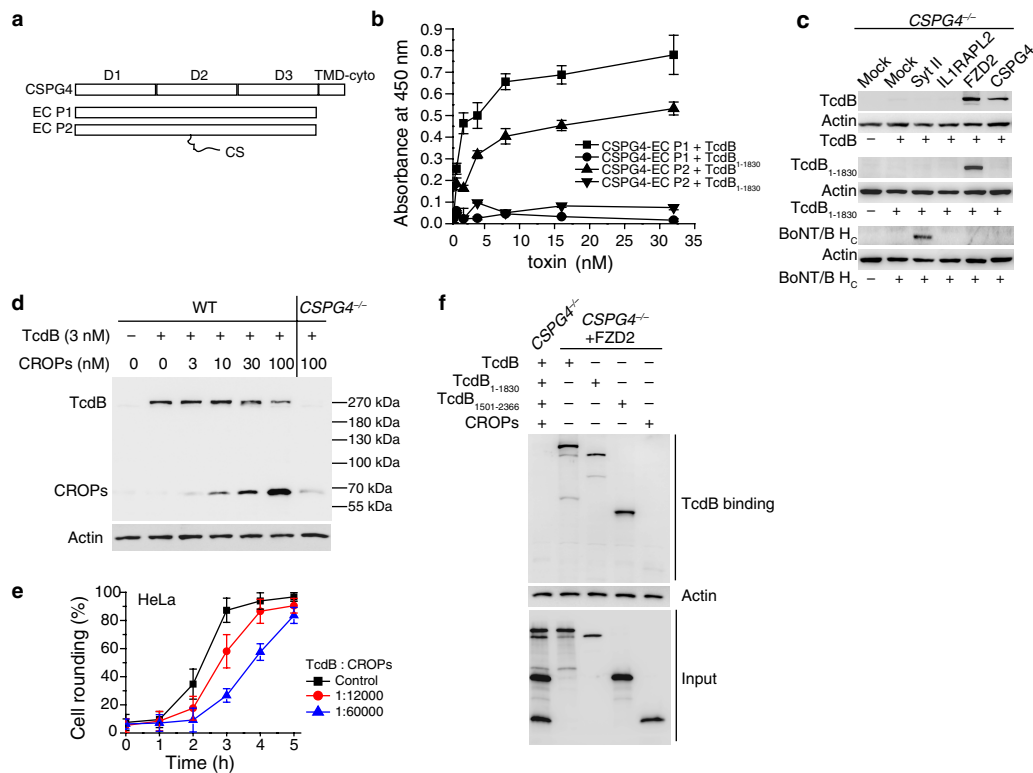
Extended Data Figure 2 | Top-ranking sgRNAs. **a**, Sequences of sgRNA were amplified by PCR after screening and subjected to NGS. The GeCKO v.2 sgRNA library is composed of two half libraries (library A and library B).

Each half library contains three unique sgRNA per gene. These two half libraries were prepared and subjected to screens independently. **b–e**, Lists of top-ranking sgRNAs. See Source Data for lists of all identified sgRNAs.



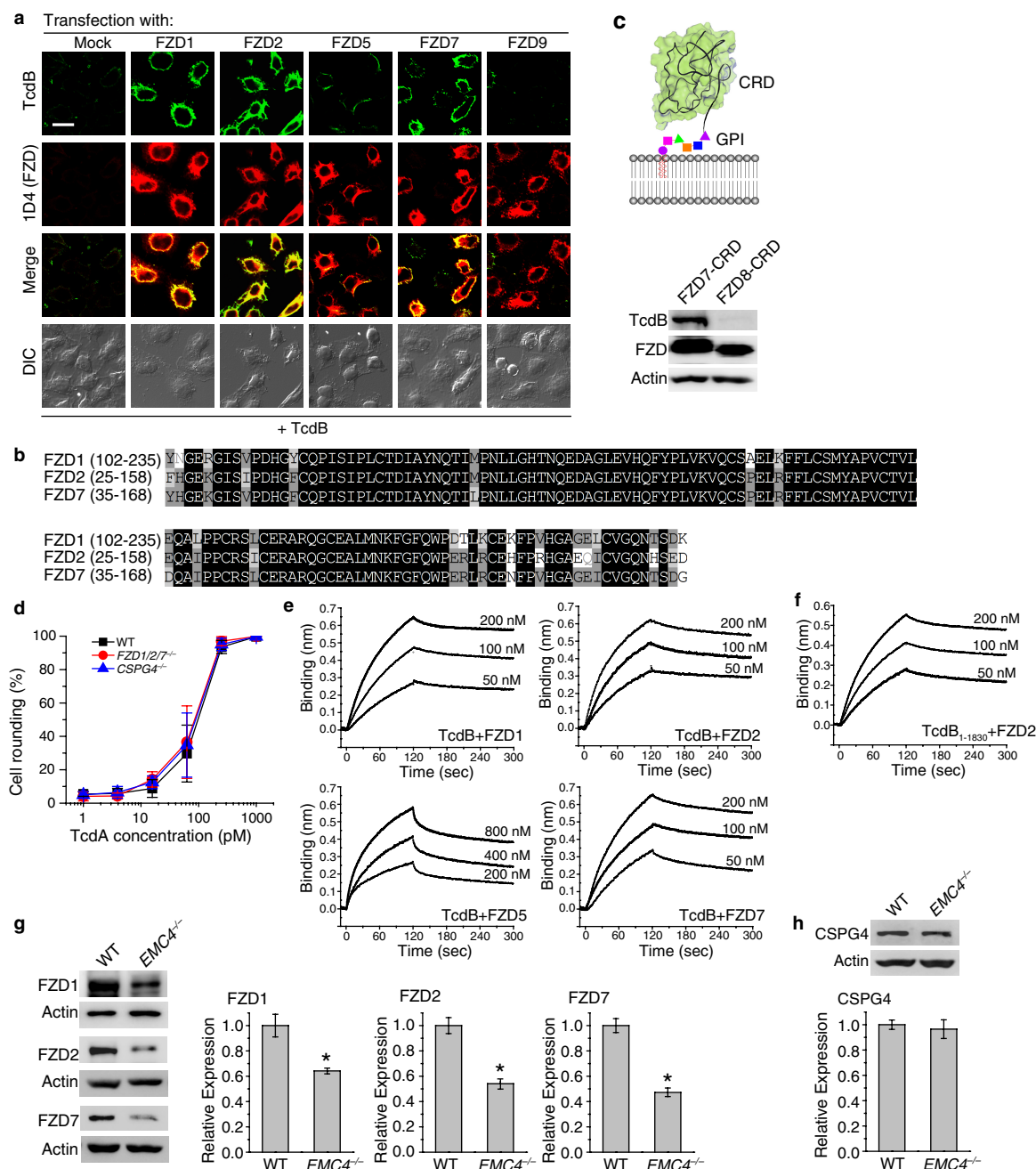
Extended Data Figure 3 | Assessing the sensitivity of HeLa knockout cells to TcdB and TcdB₁₋₁₈₃₀. **a, b,** HeLa-Cas9 cells with the indicated genes mutated via CRISPR-Cas9, as well as wild-type (WT) HeLa-Cas9 cells, were exposed to TcdB (**a**) or TcdB₁₋₁₈₃₀ (**b**) for 24 h. The percentages of rounded cells were quantified and plotted (error bars indicate mean \pm s.d., three independent experiments). **c,** HeLa knockout cells were exposed to TcdB or TcdB₁₋₁₈₃₀ for 3 h. Cell lysates were subjected

to immunoblot analysis for RAC1 and non-glucosylated (gluc.) RAC1. UGP2^{-/-} cells retained high levels of non-glucosylated RAC1 after exposure to TcdB or TcdB₁₋₁₈₃₀. CSPG4^{-/-} cells retained high levels of non-glucosylated RAC1 after exposure to TcdB. FZD2^{-/-} and EMC4^{-/-} cells showed slightly higher levels of non-glucosylated RAC1 compared to wild-type cells after exposure to TcdB₁₋₁₈₃₀. Representative blots are one from two independent experiments.



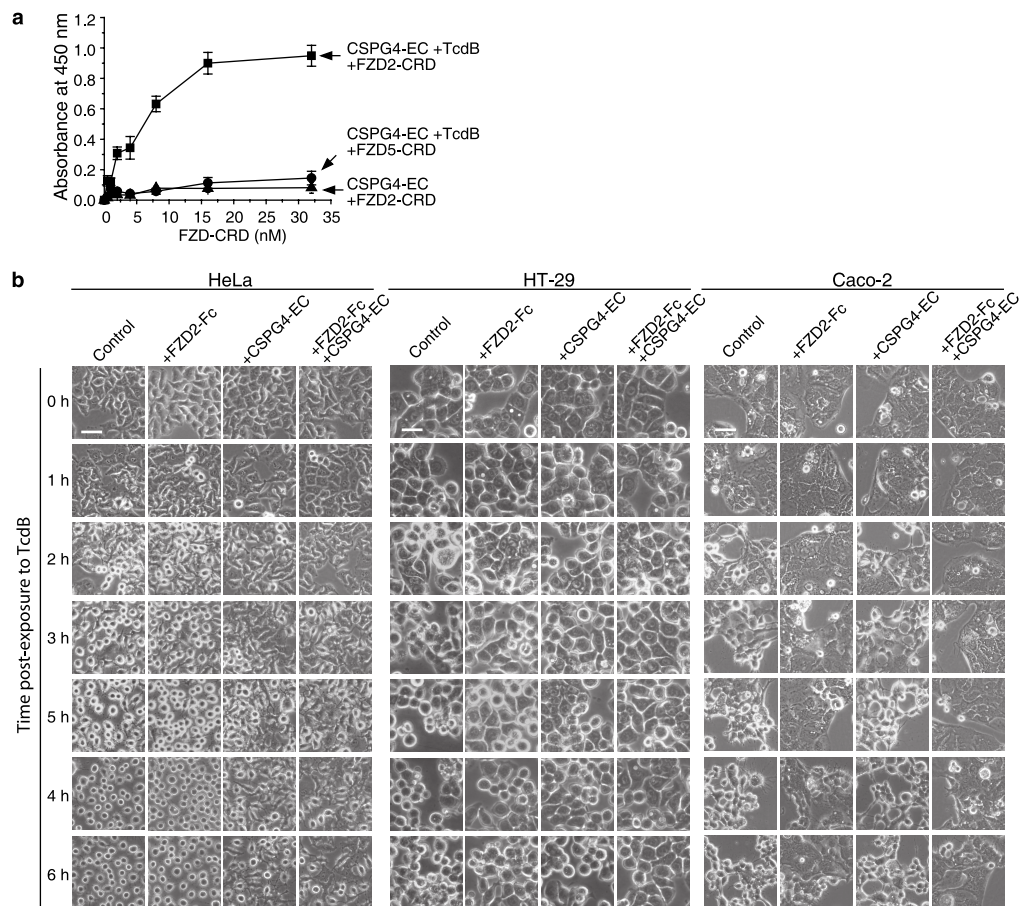
Extended Data Figure 4 | CROPs are essential for TcdB binding to CSPG4, but not required for TcdB binding to FZDs. **a**, Schematic drawings of rat CSPG4. Two pools of recombinant extracellular domain (EC) fragments were used: one that does not contain chondroitin sulfate (CS) chains (EC P1), and the other that contains CS (EC P2). TMD-cyto, transmembrane and cytoplasmic domain. **b**, TcdB, but not TcdB₁₋₁₈₃₀, binds directly to both EC P1 and EC P2 of CSPG4 in a microtitre plate-based binding assay (error bars indicate mean \pm s.d., two independent experiments). **c**, CSPG4^{-/-} cells transfected with the indicated constructs were exposed to TcdB (10 nM), TcdB₁₋₁₈₃₀ (10 nM), or the receptor-binding domain of botulinum neurotoxin B (BoNT/B H_c; 100 nM) for 10 min. Cell lysates were collected and subjected to immunoblot analysis. IL1RAPL2 and synaptotagmin II (Syt II, a receptor for BoNT/B) served as controls. Transfection of CSPG4 increased binding of TcdB, but not

TcdB₁₋₁₈₃₀, whereas transfection of FZD2 increased binding of both TcdB and TcdB₁₋₁₈₃₀. One of three independent experiments is shown. **d**, The CROP domain binds to CSPG4 on cell surfaces in a dose-dependent manner. High concentrations of recombinant CROPs reduced CSPG4-dependent binding of TcdB to cell surfaces, indicating that the CROPs can compete with TcdB for binding to CSPG4 on cell surfaces. One of three independent experiments is shown. **e**, The CROP domain reduced cytopathic toxicity of TcdB (5 pM) on wild-type (WT) HeLa cells (error bars indicate mean \pm s.d., two independent experiments). **f**, CSPG4^{-/-} cells were transfected with FZD2 and then exposed to TcdB or indicated TcdB fragments. FZD2-mediated binding of TcdB, TcdB₁₋₁₈₃₀ and TcdB₁₅₀₁₋₂₃₆₆, but not the CROPs (TcdB₁₈₃₁₋₂₃₆₆). One of three independent experiments is shown.



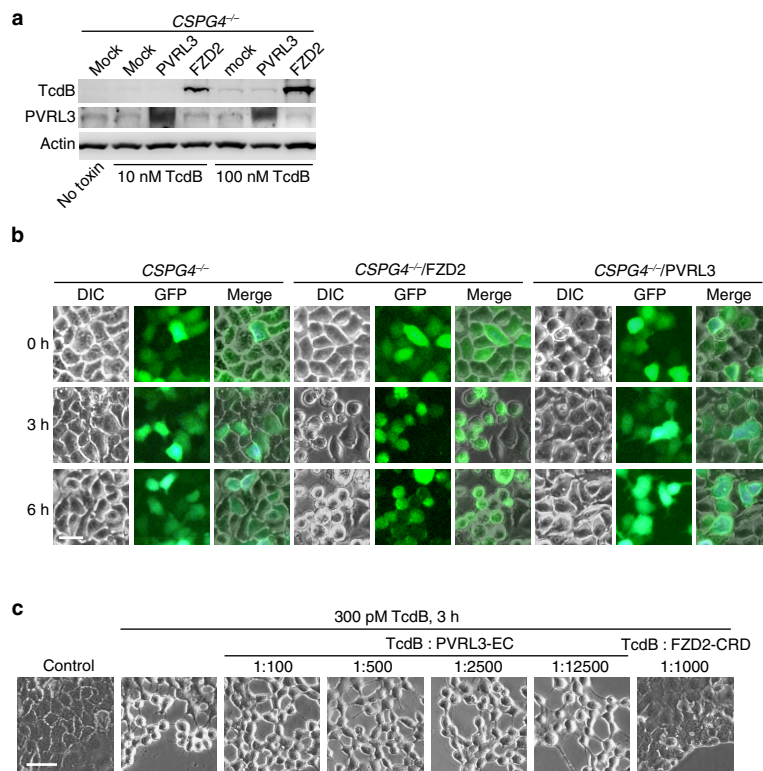
Extended Data Figure 5 | Characterizing TcdB binding to FZDs. a, *CSPG4*^{-/-} cells were transfected with 1D4-tagged FZD1, 2, 5, 7 and 9. Cells were exposed to TcdB (10 nM, 10 min), washed, fixed, permeabilized and subjected to immunostaining analysis. Scale bar, 20 μ m. One of three independent experiments is shown. **b,** The CRD domains of human FZD1 (residues 102–235), FZD2 (residues 25–158) and FZD7 (residues 35–168) were aligned using the Vector NTI software. **c,** FZD7-CRD, but not FZD8-CRD, when expressed on the surface of *CSPG4*^{-/-} cells via a GPI anchor, mediated binding of TcdB (10 nM, 10 min) to cells. One of three independent experiments is shown. **d,** Wild-type (WT) HeLa cells, *FZD1/2/7*^{-/-} cells, and *CSPG4*^{-/-} cells were exposed to TcdA and subjected to cytopathic cell-rounding assay. No reduction in sensitivity to TcdA was observed for *FZD1/2/7*^{-/-} cells or *CSPG4*^{-/-} cells, suggesting that TcdA does not use FZD1/2/7 or *CSPG4* as its receptors (error bars

indicate mean \pm s.d., two independent experiments). **e, f,** Representative binding/dissociation curves for TcdB binding to Fc-tagged CRDs of FZD1, 2, 5 and 7 (**e**), and for TcdB_{1–1830} binding to FZD2-CRD-Fc (**f**). Binding parameters are listed in Supplementary Table 3. Representative curves are from one of three independent experiments. **g,** Wild-type and *EMC4*^{-/-} cells were transfected with 1D4-tagged FZD1, 2 or 7. Cell lysates were subjected to immunoblot analysis. Expression of FZD1, 2 and 7 are reduced in *EMC4*^{-/-} cells compared to wild-type cells ($n = 6$, $*P < 0.005$, one-way ANOVA). Representative blots are from one of three independent experiments. **h,** Expression levels of *CSPG4* in *EMC4*^{-/-} cells is similar to those in wild-type cells, suggesting that EMC is not required for single-pass transmembrane proteins. One of three independent experiments is shown.



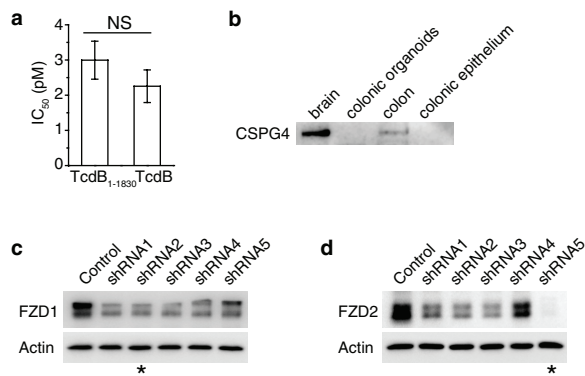
Extended Data Figure 6 | TcdB can bind to both FZD and CSPG4 simultaneously. **a**, Rat CSPG4-EC was immobilized on microtitre plates, followed by binding of TcdB, washing away unbound TcdB, and addition of FZD-CRD. FZD2-CRD binds robustly to TcdB that is pre-bound by CSPG4-EC on the microtitre plate. FZD2-CRD did not bind to CSPG4-EC without TcdB, and FZD5-CRD showed no detectable binding to CSPG4-

TcdB in this assay (error bars indicate mean \pm s.d., two independent experiments). **b**, Experiments are described in Fig. 3d on HeLa (5 pM TcdB), HT-29 (50 pM TcdB) and Caco-2 cells (150 pM TcdB). Scale bars: 50 μ m (HeLa and Caco-2) or 25 μ m (HT-29). Representative images are from one of four independent experiments.

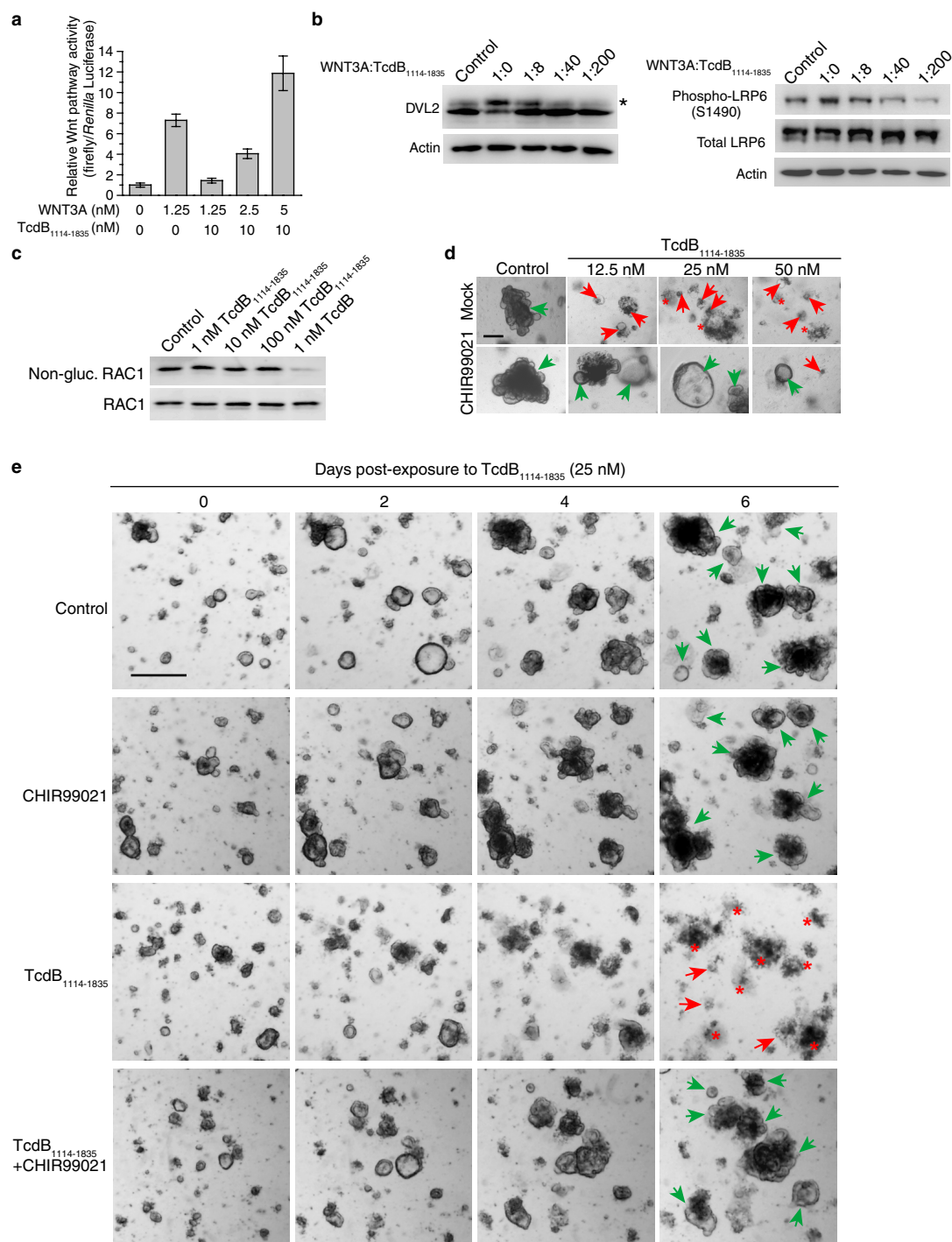


Extended Data Figure 7 | PVRL3 failed to mediate binding and entry of TcdB in HeLa and Caco-2 cells. **a**, *CSPG4*^{-/-} HeLa cells transfected with the indicated constructs were exposed to TcdB in medium for 10 min. Cell lysates were collected and subjected to immunoblot analysis. Expression of PVRL3 was confirmed using an anti-PVRL3 antibody. Transfection of FZD2, but not PVRL-3, increased binding of TcdB to *CSPG4*^{-/-} cells. One of three independent experiments is shown. **b**, Cells were challenged with TcdB (300 pM). Ectopic expression of PVRL3 failed to restore the

sensitivity of *CSPG4*^{-/-} HeLa cells towards TcdB, while expression of FZD2 restored entry of TcdB and resulted in rounding of transfected cells. Co-transfected GFP marked transfected cells. Scale bar, 50 μ m. One of three independent experiments is shown. **c**, Recombinant extracellular domain of PVRL3 (PVRL3-EC) did not reduce TcdB entry into Caco-2 cells, analysed by the cytopathic cell-rounding assay. In contrast, FZD2-CRD prevented entry of TcdB into Caco-2 cells. Scale bar, 50 μ m. One of three independent experiments is shown.

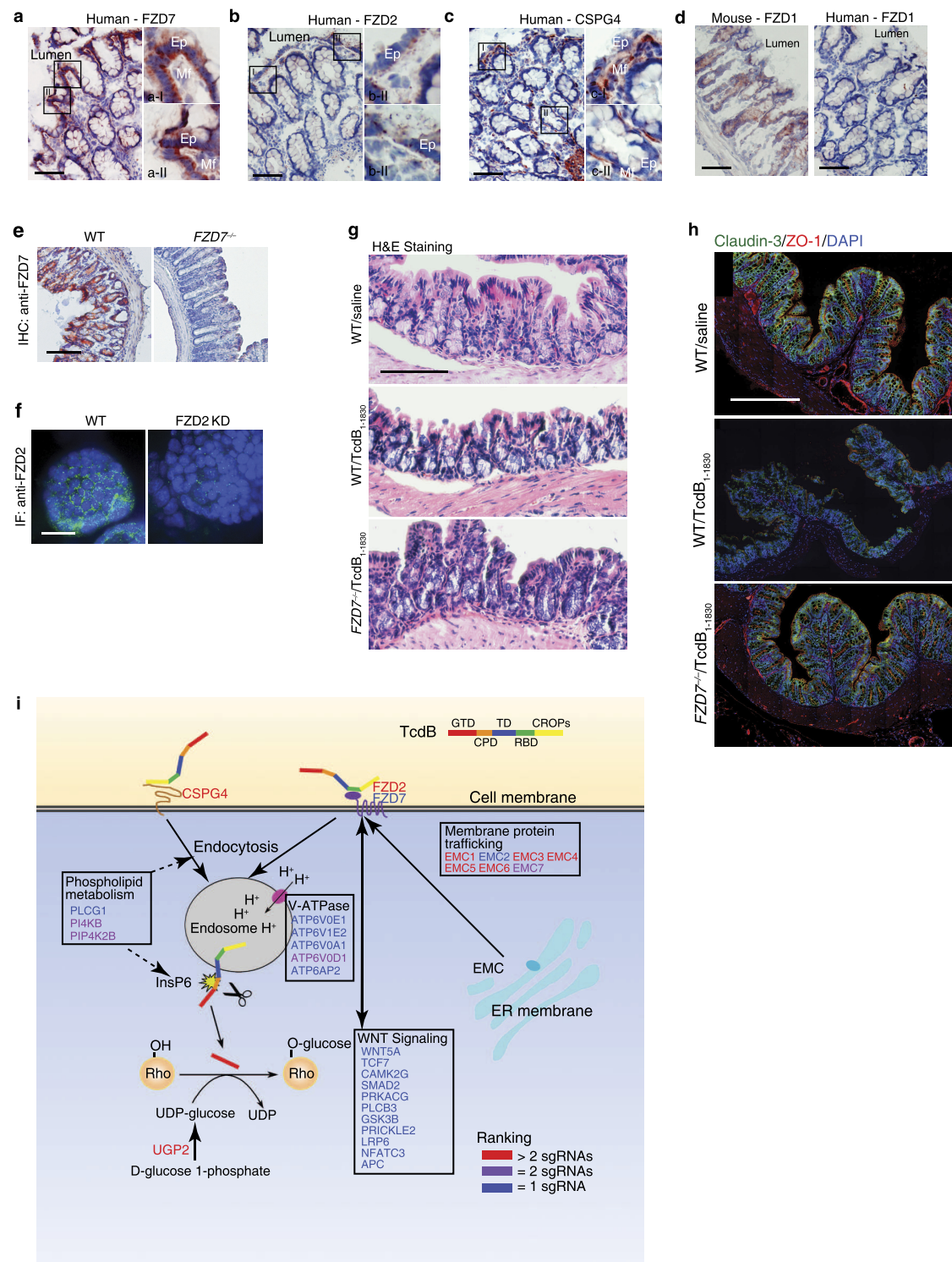


Extended Data Figure 8 | Colonic organoids showed similar levels of sensitivity to TcdB and TcdB₁₋₁₈₃₀, and validation of FZD1 and FZD2 knockdown efficiency. **a**, Colonic organoids were cultured from wild-type mice. They were exposed to a gradient of TcdB or TcdB₁₋₁₈₃₀. Viability of organoids was quantified using the MTT assay. TcdB and TcdB₁₋₁₈₃₀ showed similar IC_{50} values, suggesting that wild-type organoids are equally susceptible to TcdB and TcdB₁₋₁₈₃₀ ($n=8$, error bars indicate mean \pm s.d., two independent experiments). NS, not significant. **b**, Immunoblot analysis of CSPG4 expression in mouse brain, colonic organoids, mouse whole colon tissue, and isolated mouse colonic epithelium (200 μ g cell/tissue lysates). The colonic epithelium was isolated from colon tissues by EDTA treatment (10 mM, 2 h at 4 °C). One of three independent experiments is shown. **c**, **d**, shRNA sequences targeting FZD1 and FZD2 were validated by measuring knockdown efficiency of transfected 1D4-tagged FZD1 and FZD2 in 293T cells. shRNAs marked with asterisks (shRNA2 for FZD1 and shRNA5 for FZD2) were used to generate adenoviruses. Actin served as the loading control. One of two independent experiments is shown.



Extended Data Figure 9 | TcdB₁₁₁₄₋₁₈₃₅ inhibits Wnt signalling and induces death of colonic organoids. **a**, TcdB₁₁₁₄₋₁₈₃₅ blocked WNT3A-mediated signalling in 293T cells in a dose-dependent manner. Increasing concentrations of WNT3A restored Wnt reporter activity blocked by TcdB₁₁₁₄₋₁₈₃₅. Wnt signalling activity was analysed using the TOPFLASH/TK-*Renilla* dual luciferase reporter assay (error bars indicate mean \pm s.d., two independent experiments). We note that 1.25 nM WNT3A equals 50 ng ml⁻¹ concentration used in Fig. 4c. **b**, 293T cells in 24-well plates were exposed to WNT3A (50 ng ml⁻¹) and TcdB₁₁₁₄₋₁₈₃₅ in culture medium for 6 h. Cell lysates were harvested and subjected to immunoblot analysis for detecting phosphorylated DVL2 and LRP6. Wnt signalling activation results in phosphorylation of DVL2 and LRP6. Phosphorylated DVL2 is marked with an asterisk. One of three independent experiments is shown.

c, Mouse colonic organoids were exposed to TcdB or TcdB₁₁₁₄₋₁₈₃₅ for 12 h. Cell lysates were subjected to immunoblot analysis. No glucosylation (gluc.) of RAC1 was observed in organoids treated with TcdB₁₁₁₄₋₁₈₃₅. One of two independent experiments is shown. **d**, Colonic organoids were exposed to TcdB₁₁₁₄₋₁₈₃₅ for 72 h, with or without CHIR99021 (5 μ M). Normal organoids (green arrow), growth inhibited organoids (red arrow), and disrupted/dead organoids (asterisk) are indicated. Scale bar, 200 μ m. One of three independent experiments is shown. **e**, Time-course images of colonic organoids exposed to CHIR99021 (5 μ M), TcdB₁₁₁₄₋₁₈₃₅ (25 nM) or a combination of TcdB₁₁₁₄₋₁₈₃₅ plus CHIR99021, at 0, 2, 4 and 6 days. Normal organoids (green arrow), growth inhibited organoids (red arrow), and disrupted/dead organoids (asterisk) are indicated. Scale bar, 500 μ m. One of four independent experiments is shown.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | FZDs are receptors for TcdB in the colonic epithelium. **a–c**, Human colon cryosections were obtained from a commercial vendor and subjected to IHC analysis for detecting FZD7 (**a**), FZD2 (**b**) and CSPG4 (**c**). Ep, epithelial cells; Mf, sub-epithelial myofibroblasts. Scale bar, 50 μ m. Representative images are from one of three independent experiments. **d**, Expression of FZD1 is not detectable in mouse or human colonic tissues. One of three independent experiments is shown. **e**, FZD7 antibody labelled wild-type colonic sections, but showed no signals on colonic tissues from *FZD7*^{-/-} mice in IHC analysis, confirming the specificity of this antibody. One of three independent experiments is shown. **f**, Immunostaining of FZD2 (green) is reduced in FZD2-knockdown colonic organoids compared to control organoids, confirming the specificity of FZD2 antibody. Cell nuclei were labelled by DAPI (blue). Scale bar, 30 μ m. One of three independent experiments is shown. **g**, Experiments are described in Fig. 5g. Representative images from one of three independent experiments are shown. Scale bar, 100 μ m. **h**, Experiments were carried out as described in Fig. 5h. Low-magnification images of immunofluorescent staining of the cell–cell

junction markers claudin-3 (green) and ZO-1 (red) were stitched together to show an overview of the colon tissue. The middle panel (WT/TcdB_{1–1830}) showed disruption of the normal staining pattern for claudin-3 and ZO-1, indicating a loss of epithelial integrity, compared with both control and *FZD7*^{-/-}/TcdB_{1–1830}. Scale bar, 200 μ m. Representative images are from one of three independent experiments. **i**, A schematic overview of cellular factors identified in the CRISPR–Cas9 screen. Validated and plausible cellular factors identified in our unbiased genome-wide screens were grouped based on their presence in the same protein complexes and/or signalling pathways. The colour of the gene names reflects the number of unique sgRNAs identified. The arrows link these genes to either confirmed or plausible roles in four major steps of TcdB action: (1) receptor-mediated endocytosis; (2) low pH in the endosomes triggers conformational changes of the TD, which translocates the GTD across endosomal membranes; (3) GTD is later released via autolysis by the CPD, which is activated by the cytosolic co-factor inositol hexakisphosphate (InsP₆); (4) released GTD glucosylates small GTPases such as Rho, Rac, and CDC42.

Ultraluminous X-ray bursts in two ultracompact companions to nearby elliptical galaxies

Jimmy A. Irwin¹, W. Peter Maksym², Gregory R. Sivakoff³, Aaron J. Romanowsky^{4,5}, Dacheng Lin⁶, Tyler Speegle¹, Ian Prado¹, David Mildebrath¹, Jay Strader⁷, Jifeng Liu^{8,9} & Jon M. Miller¹⁰

A flaring X-ray source was found near the galaxy NGC 4697 (ref. 1). Two brief flares were seen, separated by four years. During each flare, the flux increased by a factor of 90 on a timescale of about one minute. There is no associated optical source at the position of the flares¹, but if the source was at the distance of NGC 4697, then the luminosities of the flares were greater than 10^{39} erg per second. Here we report the results of a search of archival X-ray data for 70 nearby galaxies looking for similar flares. We found two ultraluminous flaring sources in globular clusters or ultracompact dwarf companions of parent elliptical galaxies. One source flared once to a peak luminosity of 9×10^{40} erg per second; the other flared five times to 10^{40} erg per second. The rise times of all of the flares were less than one minute, and the flares then decayed over about an hour. When not flaring, the sources appear to be normal accreting neutron-star or black-hole X-ray binaries, but they are located in old stellar populations, unlike the magnetars, anomalous X-ray pulsars or soft γ repeaters that have repetitive flares of similar luminosities.

One flaring source (hereafter Source 1) is located at right ascension RA = 12 h 42 min 51.4 s and declination dec. = +02° 38' 35" (J2000) near the Virgo elliptical galaxy NGC 4636 (distance from Milky Way $d = 14.3$ Mpc)^{2,3}. A plot of the cumulative X-ray photon arrival time and a crude background-subtracted light curve for this source derived from an approximately 76,000-s Chandra observation taken on 2003 February 14 are shown in Fig. 1. Prior to and after the flare, the X-ray count rate of the source was $(2.1 \pm 0.2) \times 10^{-3}$ counts per second (errors given here and elsewhere, unless otherwise stated, are 1σ). This count rate corresponds to a 0.3–10-keV luminosity of $(7.9 \pm 0.8) \times 10^{38}$ erg s⁻¹ for a power-law spectral model with a best-fit photon index of $\Gamma = 1.6 \pm 0.3$ (see Methods), assuming the source is at the distance of NGC 4636. About 12,000 s into the observation, the source flared dramatically, with six photons detected in a 22-s span, leading to a conservative peak flare count rate of $0.3_{-0.1}^{+0.2}$ counts per second—an increase in emission by a factor of 70–200 over its persistent (non-flare) state. Assuming the same spectral model as in the persistent state, the flare peaks at $9_{-4}^{+6} \times 10^{40}$ erg s⁻¹. Following the initial 22-s burst, the source emitted at a less intense, but still elevated, rate for the next 1,400 s. In total, 25 photons were emitted during the flare, for an average X-ray luminosity of $(7 \pm 2) \times 10^{39}$ erg s⁻¹ and a total flare energy of $(9 \pm 2) \times 10^{42}$ erg. We assess that the probability of this burst being due to a random Poisson fluctuation of the persistent count rate is about 6×10^{-6} (see Methods). Although the photon statistics during the 25-photon burst were limited, there was no evidence that the spectrum of the source differed during the flare. There are no apparent flares in the combined 370,000 s of the other Chandra and XMM-Newton observations of NGC 4636, either before or after 2003 February 14 (see Extended Data Table 1).

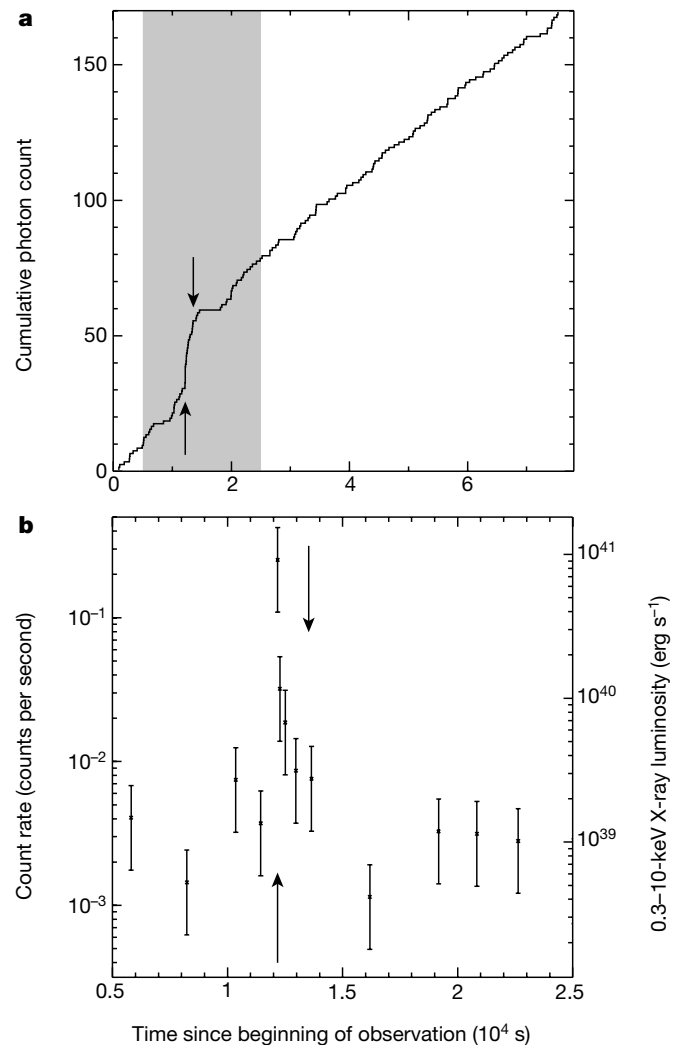


Figure 1 | Chandra cumulative X-ray photon arrival time and light curve for Source 1 in the NGC 4636 globular cluster. **a**, In total, 162 photons were detected over the approximately 76,000-s observation. The flare began after 12,000 s of observation and lasted for 1,400 s. The beginning and end of the flare are indicated by up and down arrows, respectively. **b**, Within the grey shaded region in **a** we derive the background-subtracted X-ray light curve. Each time bin contains five photons, with error bars representing the 1σ uncertainty expected from Poisson statistics.

¹Department of Physics and Astronomy, University of Alabama, Box 870324, Tuscaloosa, Alabama 35487, USA. ²Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ³Department of Physics, University of Alberta, CCIS 4-181, Edmonton, Alberta T6G 2E1, Canada. ⁴Department of Physics and Astronomy, San José State University, One Washington Square, San Jose, California 95192, USA. ⁵University of California Observatories, 1156 High Street, Santa Cruz, California 95064, USA. ⁶Space Science Center, University of New Hampshire, Durham, New Hampshire 03824, USA. ⁷Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA. ⁸Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China. ⁹College of Astronomy and Space Sciences, University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China. ¹⁰Department of Astronomy, The University of Michigan, 1085 South University Avenue, Ann Arbor, Michigan 48103, USA.

A previous study⁴ spatially associated Source 1 with a purported globular cluster of NGC 4636 that was identified through Washington C-band and Kron–Cousins R-band system CTIO Blanco Telescope imaging⁵. Although faint ($R = 23.02$), the optical source identified as CTIO ID 6444 had a $C - R = 1.94$ colour that is consistent with a globular cluster of near-solar metallicity⁵. Follow-up spectroscopic observations⁶ of a sub-sample of the globular cluster candidates in the vicinity of the globular cluster that hosts Source 1 found that 52 out of 54 (96%) of the objects with a $C - R$ colour and magnitude similar to CTIO ID 6444 were confirmed to be globular clusters of NGC 4636, with the two remaining objects identified as foreground Galactic stars. The hard X-ray spectrum of Source 1 during its persistent phase (see Methods) evidences against it being a late-type Galactic dwarf star, for which the X-ray emission tends to be quite soft. Galactic RS CVn stars exhibit hard X-ray emission in quiescence and are known to undergo X-ray flares, but these stars have much higher optical-to-X-ray flux ratios⁷ compared to Source 1. Therefore, it is highly likely that the optical counterpart of Source 1 is a globular cluster within NGC 4636. On the basis of its absolute R-band magnitude and ratio of mass to light ($M/L = 4.1$, with M in units of solar masses M_{\odot} and L in solar R-band luminosities), we estimate the globular cluster to have a mass of $3 \times 10^5 M_{\odot}$ (see Methods).

A second X-ray source located near the elliptical galaxy NGC 5128 showed similar flaring behaviour. In the 2007 March 30 Chandra observation of NGC 5128, a source at RA = 13 h 25 min 52.7 s, dec. = $-43^{\circ} 05' 46''$ (J2000; hereafter Source 2) began the observation emitting at a count rate of $(9.5 \pm 1.5) \times 10^{-4}$ counts per second. This count rate corresponds to a 0.3–10-keV luminosity of $(4.4 \pm 0.7) \times 10^{37}$ erg s⁻¹ using the best-fit $\Gamma = 1.0 \pm 0.2$ power-law photon index and a distance⁸ of 3.8 Mpc for NGC 5128. Midway through the observation, the source flared dramatically, with 10 photons detected in a 51-s time span corresponding to a conservative peak luminosity estimate of $9_{-3}^{+4} \times 10^{39}$ erg s⁻¹, after which the flare subsided. Following the flare, Source 2 returned to its pre-flare luminosity for the remainder of the observation.

Inspection of other archival Chandra and XMM-Newton data (see Extended Data Table 1) revealed four more flares of Source 2. Three were observed with Chandra on 2007 April 17, 2007 May 30 and 2009 January 4, and the fourth flare was observed with XMM-Newton on 2014 February 9. In each instance, during the initial fast (<30 s) rise of the flare the count rate increased by a factor of 200–300 over the persistent count rate to about 10^{40} erg s⁻¹, after which the flare subsided. The total flare energy of each of the five flares was approximately 10^{42} erg. The light curves for the four Chandra flares look remarkably similar, as illustrated in Fig. 2. We combined these

four light curves into a combined background-subtracted light curve (see Methods for details) in Fig. 2. Following the fast rise of the source by a factor of about 200 to a peak luminosity approaching 10^{40} erg s⁻¹, the source remained in a roughly steady ultraluminous state for approximately 200 s before decaying over a time span of around 4,000 s (Fig. 2). Fitting a power law to the combined spectra of the four Chandra flares yielded a best-fit photon index of $\Gamma = 1.2 \pm 0.3$. Therefore, much like Source 1, the spectrum of Source 2 did not change appreciably during the flare.

Source 2 has previously been identified with the object HGHH-C21 (also called GC 0320) within NGC 5128^{9–11}. With a spectroscopically determined recessional velocity¹⁰ (460 km s^{-1}) within 110 km s^{-1} of that of NGC 5128, the source is clearly at the distance of NGC 5128. This implies a projected half-light radius¹² of 7 pc. With a velocity dispersion of 20 km s^{-1} and an inferred stellar mass¹² of $3.1 \times 10^6 M_{\odot}$, the optical counterpart is either a massive globular cluster or, given its unusual elongated shape, more likely an ultracompact dwarf companion galaxy of NGC 5128.

It is unlikely that the flaring and the steady emission in both sources are attributable to two unrelated sources in the same host. Because our flare search technique would have found these flares had they been detected by their flare photons alone, we can calculate the probability that these globular clusters would have also hosted steady X-ray emission more luminous than the persistent emission in each globular cluster (see Methods). The globular cluster in Source 1 has a <0.3% probability of having an X-ray source with a luminosity of more than 8×10^{38} erg s⁻¹; the globular cluster or ultracompact dwarf in Source 2 has a <9% probability of having an X-ray source with a luminosity of more than 4×10^{37} erg s⁻¹. Multiplying these probabilities leads to only a <0.02% chance that both flares are unrelated to the steady emission. In the unlikely event that the flares are distinct sources from the persistent sources, the flaring sources must be flaring by more than two orders of magnitude over whatever their true non-flare luminosities are.

Summing up all the available archival Chandra and XMM-Newton data (but omitting the Chandra High Resolution Camera (HRC) and transmission grating exposures, which are not sensitive enough to detect a flare of similar intensity to that seen in the Advanced CCD Imaging Spectrometer (ACIS) observations) allows us to constrain the duty cycle and recurrence rate of the flares. Source 2 flared five times for a total combined flare time of about 20,000 s in a total observation time of 7.9×10^5 s, yielding one flare every approximately 1.8 days and a duty cycle of about 2.5%. Source 1 flared once for 1,400 s in a total observation time of 370,000 s. This single flare implies a recurrence timescale of >4 days and duty cycle of <0.4%.

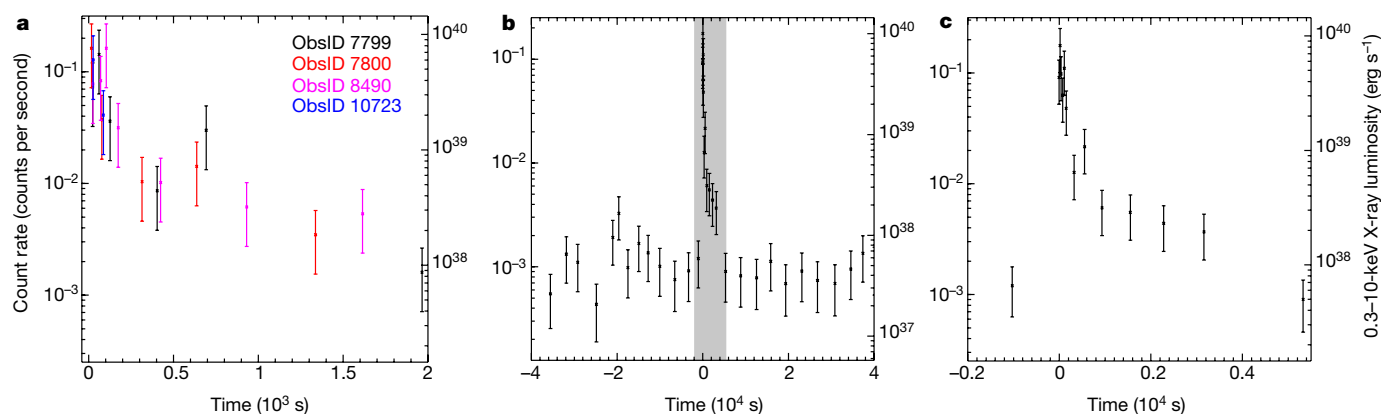


Figure 2 | Individual and combined background-subtracted X-ray light curves for Source 2 in the NGC 5128 globular cluster or ultracompact dwarf. **a**, The X-ray light curves for the four Chandra flares show similar behaviour. Each time bin contains five photons. **b**, The combined light curve of the four flares illustrates the fast rise and slow decay of the flares.

Each time bin contains ten photons. The time is given relative to the beginning of the flare. **c**, Zooming in on the grey shaded region in **b** reveals that the luminosity during the flare rose quickly and remained steady in an ultraluminous state for approximately 200 s before decaying back to its persistent level after about 1 h. All error bars represent 1σ uncertainties.

In terms of energetics, variability and survivability, only short- and intermediate-duration soft gamma repeaters (SGRs)¹³ and their cousins the anomalous X-ray pulsars (AXPs)¹⁴ are comparable to the sources discussed here. However, SGRs and AXPs are believed to be very young and highly magnetized neutron stars, which are not likely to be found in an old stellar population such as a globular cluster or red ultracompact dwarf galaxy. Our sources are also unlike SGRs and AXPs in that SGR/AXP flares of this magnitude last only a few to a few tens of seconds^{15,16} without an hour-long decay as seen in our sources. Our sources are also unlikely to be type-II X-ray bursts of neutron stars, which are believed to result from rapid spasmodic accretion onto the neutron star. In addition to having flare-to-pre-flare luminosity ratios of only 10–20, the only type-II burst to reach 10^{40} erg s^{−1} (GRO J1744–28, the Bursting Pulsar) exhibits several sub-minute flares per day when flaring, with total flare energies per burst that are much lower than our sources and different timing properties from our sources¹⁷. Furthermore, the quiescent X-ray luminosity of the Bursting Pulsar is 4–5 orders of magnitude fainter than the long-term luminosities of our sources. Qualitatively, the fast rise and slower decay of Source 2 (Fig. 2) resembles that of type-I bursts from Galactic neutron stars, which typically peak near the Eddington limit of a neutron star. However, the peak luminosities from Sources 1 and 2 are 1–2 orders of magnitude greater than the type-I limit for even helium accretion and last more than an order of magnitude longer. Rare superbursts from Galactic neutron stars have been known to last for an hour^{18,19}, but have peak luminosities well below those of our sources. Other X-ray flares of unknown source^{20–22} appear to be one-time transient events, indicating that they were (probably) cataclysmic events with no post-flare emission, unlike our sources.

We investigated the light curves of several thousand X-ray point sources within 70 Chandra observations of nearby galaxies and found only the two examples presented here. It would appear that the Milky Way has no analogues to our sources. This is not surprising given the small number (about 40)²³ of X-ray sources in the Milky Way that are brighter than 10^{37} erg s^{−1}, the lack of X-ray binaries that are more luminous than 10^{38} erg s^{−1} in Galactic globular clusters, and the rarity of burst sources in the extragalactic sample. The nature of Sources 1 and 2 remains uncertain. The increased emission during the burst might result from a narrow cone of beamed emission that crosses our line of sight every few days. However, it is unclear how a pulsed beam would lead to the distinctly asymmetric fast rise and slower decay profile. Alternatively, the flare might represent a period of rapid, highly super-Eddington accretion onto a neutron star or stellar-mass black hole, perhaps during the periastron passage of a donor companion star in an eccentric orbit around a compact object. Such an explanation has been suggested to explain observed (albeit neutron-star Eddington-limited) flares in galaxies^{1,24}. Finally, the high X-ray luminosity during the peak of the flare might represent accretion onto an intermediate-mass black hole. If the flares are Eddington-limited, then black hole masses of $800M_{\odot}$ and $80M_{\odot}$ are implied for Sources 1 and 2, respectively, assuming a bolometric correction of 1.1 appropriate for a 2-keV-disk blackbody-temperature spectral model. The fast rise times constrain the maximum mass of a putative black hole, because the rise time cannot be shorter than the travel time of light across the innermost stable circular orbit of the black hole. For both sources, the fastest rise happened over a 22-s period, implying an upper limit on the mass of a maximally rotating black hole of $2 \times 10^6 M_{\odot}$. A black hole in this mass range is a particularly intriguing explanation for Source 2, if indeed its host is the stripped core of a dwarf galaxy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 June; accepted 31 August 2016.

1. Sivakoff, G. R., Sarazin, C. L. & Jordán, A. Luminous X-ray flares from low-mass X-ray binary candidates in the early-type galaxy NGC 4697. *Astrophys. J.* **624**, L17–L20 (2005).
2. Tonry, J. L. *et al.* The SBF survey of galaxy distances. IV. SBF magnitudes, colors, and distances. *Astrophys. J.* **546**, 681–693 (2001).
3. Mei, S. *et al.* The ACS Virgo Cluster survey. XIII. SBF distance catalog and the three-dimensional structure of the Virgo Cluster. *Astrophys. J.* **655**, 144–162 (2007).
4. Posson-Brown, J., Raychaudhury, S., Forman, W., Donnelly, R. H. & Jones, C. Chandra observations of the X-Ray point source population in NGC 4636. *Astrophys. J.* **695**, 1094–1110 (2009).
5. Dirsch, B., Schuberth, Y. & Richtler, T. A wide-field photometric study of the globular cluster system of NGC 4636. *Astron. Astrophys.* **433**, 43–56 (2005).
6. Schuberth, Y. *et al.* Dynamics of the NGC 4636 globular cluster system. An extremely dark matter dominated galaxy? *Astron. Astrophys.* **459**, 391–406 (2006).
7. Pandey, J. C. & Singh, K. P. A study of X-ray flares – II. RS CVn-type binaries. *Mon. Not. R. Astron. Soc.* **419**, 1219–1237 (2012).
8. Harris, G. L. H., Rejkuba, M. & Harris, W. E. The distance to NGC 5128 (Centaurus A). *Publ. Astron. Soc. Aust.* **27**, 457–462 (2010).
9. Harris, G. L. H., Geisler, D., Harris, H. C. & Hesser, J. E. Metal abundances from Washington photometry of globular clusters in NGC 5128. *Astron. J.* **104**, 613–626 (1992).
10. Woodley, K. A. *et al.* The kinematics and dynamics of the globular clusters and planetary nebulae of NGC 5128. *Astron. J.* **134**, 494–510 (2007).
11. Woodley, K. A. *et al.* Globular clusters and X-Ray point sources in Centaurus A (NGC 5128). *Astrophys. J.* **682**, 199–211 (2008).
12. Mieske, S. *et al.* On central black holes in ultra-compact dwarf galaxies. *Astron. Astrophys.* **558**, A14 (2013).
13. Kouveliotou, C. *et al.* An X-ray pulsar with a superstrong magnetic field in the soft γ -ray repeater SGR1806–20. *Nature* **393**, 235–237 (1998).
14. Mereghetti, S. & Stella, L. The very low mass X-ray binary pulsars: a new class of sources? *Astrophys. J.* **442**, L17–L20 (1995).
15. Olive, J.-F. *et al.* Time-resolved X-Ray spectral modeling of an intermediate burst from SGR 1900+14 observed by *HETE-2* FREGATE and *WXM*. *Astrophys. J.* **616**, 1148–1158 (2004).
16. Kozlova, A. V. R. *et al.* The first observation of an intermediate flare from SGR 1935+2154. *Mon. Not. R. Astron. Soc.* **460**, 2008–2014 (2016).
17. Younes, G. *et al.* Simultaneous NuSTAR/Chandra observations of the bursting pulsar GRO J1744–28 during its third reactivation. *Astrophys. J.* **804**, 43 (2015).
18. Cornille, R., Heise, J., Kuulkers, E., Verbunt, F. & in 't Zand, J. J. M. The longest thermonuclear X-ray burst ever observed? A BeppoSAX Wide Field Camera observation of 4U 1735–44. *Astron. Astrophys.* **357**, L21–L24 (2000).
19. Strohmayer, T. E. & Beown, E. F. A remarkable 3 hour thermonuclear burst from 4U 1820–30. *Astrophys. J.* **566**, 1045–1059 (2002).
20. Jonker, P. G. *et al.* Discovery of a new kind of explosive X-Ray transient near M86. *Astrophys. J.* **779**, 14 (2013).
21. Luo, B., Brandt, W. N. & Bauer, F. Discovery of a fast X-ray transient in the Chandra Deep Field-South survey. *Astron. Telegr.* 6541 (2014).
22. Glennie, A., Jonker, P. G., Fender, R. P., Nagayama, T. & Pretorius, M. L. Two fast X-ray transients in archival Chandra data. *Mon. Not. R. Astron. Soc.* **450**, 3765–3770 (2015).
23. Grimm, H.-J., Gilfanov, M. & Sunyaev, R. The Milky Way in X-rays for an outside observer. Log(*N*)-Log(*S*) and luminosity function of X-ray binaries from RXTE/ASM data. *Astron. Astrophys.* **391**, 923–944 (2002).
24. Maccarone, T. J. An explanation for long flares from extragalactic globular cluster X-ray sources. *Mon. Not. R. Astron. Soc.* **364**, 971–976 (2005).

Acknowledgements We thank T. Richtler for discussions. J.A.I. was supported by Chandra grant AR6-17010X and NASA ADAP grant NNX10AE15G. G.R.S. acknowledges the support of an NSERC Discovery Grant. A.J.R. was supported by the National Science Foundation grant AST-1515084. J.S. acknowledges support from NSF grants AST-1308124 and AST-1514763 and the Packard Foundation.

Author Contributions J.A.I. led the Chandra data reduction and analysis, with contributions from W.P.M. for the XMM-Newton data reduction and analysis. T.S., I.P. and D.M. conducted the Chandra galaxy survey that yielded the two flare sources, with oversight from J.A.I. G.R.S., A.J.R., D.L., J.S., J.L. and J.M.M. contributed to the discussion and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.I. (jairwin@ua.edu).

METHODS

Flare search technique. We searched for flares from all point sources found in 70 Chandra observations of nearby luminous early-type galaxies. The *evt2* files were downloaded from the Chandra archive, and the source-detection routine *wavdetect* in the Chandra Interactive Analysis of Observations (CIAO) package suite was used on the image files to create a list of sources detected at the $>3\sigma$ level. Our script then extracted the time-ordered photon arrival times for each source found by *wavdetect*. Next, our routines scanned the photon event list and searched for bursts by finding the time difference between each photon and the photon three photons forward in time from it (that is, a 4-photon burst). They then calculated the Poisson probability of detecting that many photons over that time interval given the overall count rate of the source over the entire observation and the number of 4-photon burst trials present over the epoch. This was repeated for 5-photon bursts, 6-photon bursts, 7-photon bursts and so on, up to 20-photon bursts. If the probability of a burst of that magnitude from Poisson fluctuations was below our fiducial value (1 in 10^6) and the count rate during the N -photon burst was more than ten times the average count rate of the source over the observation, then the source was marked for further study. Note that our technique is more sophisticated than a simple Kolmogorov–Smirnov test on this distribution. Our technique found several (previously known and unknown) flares from Milky Way M dwarf stars, which were removed from consideration. Among the 7,745 sources detected in the 70 observations, Source 1 and Source 2 were the only non-Galactic sources not previously detected (which appear to be transients or one-time events^{20–22}), for which the random Poisson fluctuation probability was less than 10^{-6} and for which peak-to-persistent count ratio exceeded ten (to exclude somewhat variable, but non-flaring sources).

Chandra and XMM-Newton data reduction. Those Chandra observations containing flares were then analysed further using CIAO 4.7 with CALDB version 4.6.9. The sources exhibited flaring in only ObsID 3926 for Source 1 and in ObsIDs 7799, 7800, 8490 and 10723 for Source 2. The remaining 3 and 36 ObsIDs for Sources 1 and 2, respectively, showed no flaring activity, did not have the source in the field of view of the detector, or the data were taken with a lower sensitivity detector (Chandra HRC/LETG/HETG) for which a flare of comparable intensity would not have been detected. Extended Data Table 1 lists all of the searched Chandra and XMM-Newton observations of NGC 4636 and NGC 5128. The luminosities of the sources in the non-flare observations were consistent with the persistent luminosities of the sources during the flare observations. All flares occurred in ACIS-I pointings. The event lists were reprocessed using the latest calibration files at the time of analysis with the CIAO tool *chandra_repro*. None of the Chandra observations had any background flaring time intervals that were significant enough to warrant their removal considering that we are interested in point sources. Energy channels below 0.3 keV and above 6.0 keV were ignored. Sources 1 and 2 were both located at least $4.8'$ from the ACIS-I aim point in all of the observations, so we used the CIAO tool *psfsize_srcs* to determine the extraction radius for each observation that enclosed 90% of the source photons at an energy of 2.3 keV. All subsequent count rates and 0.3–10-keV X-ray luminosities were corrected for these point spread function (PSF) losses. Because each flare occurred at a large off-axis angle from the aim point, the photons were spread over a large PSF. Consequently, pile-up effects were negligible even at the peak of each flare. We do not believe that there is any way for the flares to be an instrumental effect such as pixel flaring or cosmic ray afterglows, for which each recorded event during the flare would occur in a single detector pixel. Inspection of all of the flare observations in detector coordinates revealed that the photons were not concentrated in a single detector pixel, but were spread out in detector space in accordance with the dither pattern of Chandra as would be expected from astrophysical photons. Furthermore, the photon energies of cosmic ray afterglows decrease with each successive photon, which is not the case for the photons occurring during the flare.

Although we did not conduct a survey of galaxies observed with XMM-Newton, we did utilize archival XMM-Newton data to search for additional flares from Sources 1 and 2. The XMM-Newton observations for Source 1 did not reveal any detectable flaring behaviour, but the 2014 February 9 observation (ObsID 0724060801) of Source 2 revealed a fifth flare for this source that was detected in the MOS1 and MOS2 detectors separately. To analyse the data, we used the 2014 November 4 release of the XMM-Newton Science Analysis Software (SAS), and the data were processed with the tool *emproc*, which filtered the data for the standard MOS event grades. Source 2 was observed only with the two MOS instruments, because it was not in the field of view of the pn camera (the observations used a restricted window owing to the high count rate of the central active galactic nucleus in NGC 5128). Periods of high background at the end of the observation were removed.

Plots of cumulative photon arrival time for Sources 1 and 2. Each X-ray photon collected by Chandra or XMM-Newton is tagged with a position, energy and time

of arrival, allowing a photon-by-photon account of each X-ray source at a time resolution set by the read-out time for the detector (3.1 s for Chandra ACIS-I and 2.6 s for XMM-Newton EPIC MOS). Plots of the cumulative photon arrival time are a simple way to observe time variability over the course of the observation. Whereas a source with constant flux will yield a cumulative photon arrival time plot with a constant slope, a flare will appear as a nearly vertical rise in the plot as photons stream in over a short period of time. Figure 1 shows the plot of cumulative photon arrival time for Source 1, illustrating the onset of the flare around 12,000 s after the beginning of the observation.

The plot of the cumulative photon arrival time from each of the five flares from Source 2 is shown in Extended Data Fig. 1. The beginning of the flare is evident in each plot. The final Chandra flare (ObsID 10723) occurred just at the end of this short observation. The persistent count rate within the $10''$ source-extraction region of the XMM-Newton observation is compromised by background, but the onset of the flare about 16,000 s after the beginning of the observation is evident.

Peak flare rate and the statistical significance of the flares. We estimated the peak flare rate of Source 1 on the basis of the arrival times of the first six photons of the flare, which arrived over a 22-s period. Given the uncertainty in when the peak ended, we neglect the sixth photon of the flare to conservatively estimate a count rate of $0.25^{+0.17}_{-0.11}$ counts per second (1σ uncertainty) after correcting for the 10% of emission that is expected to be scattered out of the source extraction region owing to PSF losses. Background was negligible during the flare and accounted for only 7% of the emission inside the source-extraction region during persistent times.

Because Source 1 flared only once, it is necessary to accurately determine the number of independent trials that were contained in the sources searched within our sample of galaxies to determine the likelihood that the flare could result from a random fluctuation in the persistent count rate. The two sources discussed here were found as part of a 70-observation sample observed with Chandra and composed primarily of large elliptical galaxies at distances of <20 Mpc, with a majority of the galaxies residing in Virgo or Fornax. Within these 70 observations, 7,745 sources were detected yielding a total of 8.5×10^5 photons. This is equivalent to 1.7×10^5 independent 5-photon groupings. Statistically, the chance of detecting five or more photons in 22 s for a source that normally emits at 1.9×10^{-3} counts per second (the count rate in the persistent state before correcting for PSF losses) is 1.0×10^{-9} . With 1.7×10^5 independent trials throughout our sample, the chance of finding a single 5-photon burst for the Source 1 flare is 1.7×10^{-4} . Searching over multiple photon-burst scales increases the odds of finding a chance statistical fluctuation. A previous study¹ that reports a similar calculation gives this correction factor to be approximately 2.5 using Monte Carlo simulations; applying that correction here leads to a false detection probability of 4.3×10^{-4} . A similar exercise for the entire burst (25 photons in 1,400 s) leads to a chance fluctuation probability of 6.4×10^{-6} .

Similar calculations can be performed for each flare detected in Source 2 by Chandra observations. In the four cases, the flare at its peak was detected using 9 photons in 51 s, 6 photons in 22 s, 7 photons in 22 s and 6 photons in 37 s. Given the persistent count rates in each observation, and correcting appropriately for both the number of independent 9-photon, 6-photon, and 7-photon trials (that is, scaling appropriately from the 1.7×10^5 independent 5-photon trials) and for the multi-burst search correction factor of 2.5, we calculate probabilities of 1.4×10^{-5} , 7.1×10^{-7} , 1.2×10^{-6} and 9.0×10^{-4} of a false flare detection for each of the four flares. Because we considered XMM-Newton data only after having detected the flares in the Chandra data, the probability that the flare observed with XMM-Newton was falsely detected is 5.0×10^{-8} given the 113,000 s of total exposure on this source. When combined, this gives a probability that all the flares were falsely detected of 5.4×10^{-28} .

X-ray light curves. Owing to the limited photon statistics for the flare in Source 1, only a crude X-ray light curve was obtained by binning photons in groups of five and determining the count rate over which the five photons were collected (Fig. 1). The four individual 5-photon-bin Chandra light curves for Source 2 showed similar timing behaviour (Fig. 2), which gave us confidence to combine them into one light curve. For each flare, we determined the average arrival time of the first three photons of the flare and set this to 'time zero'. Thus, photons before the flare were assigned a negative time value. The four photon lists were then combined at 'time zero' to provide a combined photon list. Photons were then binned in groups of ten to calculate count rates during the time period over which the ten photons were collected. The count rates were divided by four to give the average count rate per time bin per flare. Because the fourth Chandra epoch (ObsID 10723) was very short and does not extend from $-40,000$ s to $40,000$ s from the start of the flare, we corrected the count rate accordingly to account for the temporal coverage of this epoch. The count rates were corrected for the loss of photons outside the extraction region due to the PSF, and for the expected background (although negligible during the flare, this accounted for 14% of the

emission during persistent periods). The combined light curve for Source 2 is shown in Fig. 2. A sharp rise at the beginning of the flare was followed by a flat ultraluminous state for about 200 s. The improvement in statistics by combining the four light curves traces the duration of the decay in flux out to about 4,000 s. Following the flare, the count rate of the source was remarkably consistent with the pre-flare count rate.

Spectral fitting and source luminosities. For Source 1, we extracted a combined spectrum during the pre- and post-flare period using the CIAO tool *specextract*. Background was collected from a source-free region surrounding our source. Using XSPECv12.8, a power-law model absorbed by the Galactic column density in the direction of NGC 4636 ($N_{\text{H}} = 1.8 \times 10^{20} \text{ cm}^{-2}$)²⁵ using the *tbabs* absorption model was used to fit the background-subtracted spectrum. Only energy channels over the range 0.5–6.0 keV were considered in the fit. The spectrum was grouped to contain at least one count per channel and the C-statistic was used in the fit. A best-fit power-law photon index of 1.6 ± 0.3 (90% uncertainty) was found. This fit implies an unabsorbed luminosity of $(7.9 \pm 0.8) \times 10^{38} \text{ erg s}^{-1}$ during the persistent state (all luminosities reported below have also been corrected for absorption). Because the flare period contained only 25 photons, the flare spectrum was poorly constrained ($\Gamma = 1.6 \pm 0.7$). This led to a peak luminosity during the first 22 s of the flare of $9_{-4}^{+6} \times 10^{40} \text{ erg s}^{-1}$, a factor of about 120 times greater than during persistent periods combined. Freeing the absorption did not substantially change the fit. Fitting the flare with a disk blackbody model gave a slightly worse fit with $kT_{\text{blackbody}} = 1.3_{-0.5}^{+2.0} \text{ keV}$ and a luminosity 30% less than that derived from the power-law fit.

For Source 2, we combined the spectra from the flare periods of the four Chandra observations into one spectrum using *specextract*. The same was done for the pre- and post-flare periods combined. The best-fit power-law photon indices for persistent and flare periods assuming a Galactic column density in the direction of NGC 5128 ($8.6 \times 10^{20} \text{ cm}^{-2}$)²⁵ were 1.0 ± 0.2 and 1.2 ± 0.3 (90% uncertainty), respectively. Again, this indicates no significant change in the spectral shape during the flare. These spectral models implied persistent and peak flare luminosities of $(4.4 \pm 0.3) \times 10^{37} \text{ erg s}^{-1}$ and $8.1_{-2.5}^{+3.5} \times 10^{39} \text{ erg s}^{-1}$, respectively—an increase of about 200 in less than a minute. When we split the flare period into the flat ultraluminous (first 200 s) and decay (200–4,000 s) times, we also found no significant spectral evolution. We allowed the Galactic column density N_{H} to vary in the fits and found a softer photon index ($\Gamma = 1.6 \pm 0.6$ in the persistent state and $\Gamma = 1.3 \pm 0.7$ during the flare) with $N_{\text{H}} = 6_{-4}^{+5} \times 10^{21} \text{ cm}^{-2}$ for the persistent state and unconstrained below $5 \times 10^{21} \text{ cm}^{-2}$ during the flare (90% uncertainties for two interesting parameters). In both instances, freeing the absorption changed the unabsorbed X-ray luminosity by only <10%. The source does not reside in the dust lane of NGC 5128, so this excess absorption, if real, might be intrinsic to the source. We also fitted the flare spectrum with a disk blackbody model with fixed N_{H} at the Galactic value and found a best-fit temperature of $2.2_{-0.6}^{+1.7} \text{ keV}$, with a comparable goodness-of-fit to that of the power-law model and a luminosity 20% below that derived from the power-law fit.

For the XMM-Newton observation, the spectrum and response files were generated using the standard SAS tasks *evselect*, *backscale*, *arfgen* and *rmfgen*. Because the count rate during the pre- and post-flare time period is dominated by background (owing to the much larger extraction region compared to Chandra and to higher background rates), we did not extract a spectrum for the persistent period. We extracted the background-subtracted flare spectrum in a $30''$ region around the source and fitted it with the absorbed power law described above for Chandra observations. The slope of the power law was poorly constrained ($\Gamma = 1.5 \pm 0.5$) owing to the low number of photons detected in the flare, but the slope was consistent with the fit from the co-added Chandra spectrum. The peak luminosity of the flare was $1.6_{-0.7}^{+1.1} \times 10^{40} \text{ erg s}^{-1}$, again consistent with the Chandra flares.

Probability of the flare and persistent emission being from two different sources. We have assumed that the persistent and flare emission emanate from a single source within the globular cluster hosts of Sources 1 and 2, but it is possible that two separate sources in the same cluster are responsible for the emission. The probability that a globular cluster hosts an X-ray binary of a particular X-ray luminosity depends on the luminosity of the source²⁶ and the properties of the globular cluster, such as its mass, concentration and metal abundance^{27,28}. From previous work²⁷, the number of X-ray sources more luminous than $3.2 \times 10^{38} \text{ erg s}^{-1}$ in a globular cluster that has a mass M , stellar encounter rate Γ_{h} , half-light radius r_{h} , and cluster metallicity Z is

$$0.041 \left(\frac{\Gamma_{\text{h}}}{10^7} \right)^{0.82 \pm 0.05} \left(\frac{Z}{Z_{\odot}} \right)^{0.39 \pm 0.07}$$

where

$$\Gamma_{\text{h}} = \left(\frac{M}{M_{\odot}} \frac{1}{2\pi} \right)^{3/2} \left(\frac{r_{\text{h}}}{1 \text{ pc}} \right)^{-5/2}$$

The globular cluster hosting Source 1 has photometry in Kron–Cousins R-band and Washington C-band filters; $R = 23.02$ and $C - R = 1.94$. This colour corresponds to a photometrically derived metallicity of $Z/H = -0.08 \text{ dex}$ ($Z = 0.8Z_{\odot}$)²⁹. Using a single population model³⁰ given a Kroupa initial mass function, a 13-Gyr age, $Z/H = -0.08 \text{ dex}$ and $M/L = 4.1$ in the R-band are expected for this cluster. Given the distance to NGC 4636 ($d = 14.3 \text{ Mpc}$), the R-band M/L referenced above, and a solar R-band magnitude $M_{\text{R}\odot} = 4.42$, we estimate a globular cluster mass of $3.0 \times 10^5 M_{\odot}$. Because we do not have a size measurement for this globular cluster, we conservatively estimate a minimum size of 1.5 pc, which is the 3σ lower limit based on a survey of globular clusters in the Virgo cluster³¹. With these values, we estimate that the globular cluster is expected to have 0.017 X-ray binaries with luminosities of more than $3.2 \times 10^{38} \text{ erg s}^{-1}$. To determine the number of X-ray binaries that are expected above the observed persistent X-ray luminosity of Source 1, we apply the X-ray luminosity function in globular clusters found in a previous study²⁶, which predicts that the Source 1 persistent luminosity ($8 \times 10^{38} \text{ erg s}^{-1}$) is ten times less likely to be found in a globular cluster than a $3.2 \times 10^{38} \text{ erg s}^{-1}$ source. This leads to an estimate of 0.0017 X-ray sources equal to or more luminous than Source 1. Therefore, after having found a flaring source, the probability that the persistent emission comes from a different X-ray binary in this cluster is <0.17%. If we conservatively assume that the predicted number of X-ray binaries could be 50% higher (approximately convolving all of the uncertainty sources), then the probability is <0.24%.

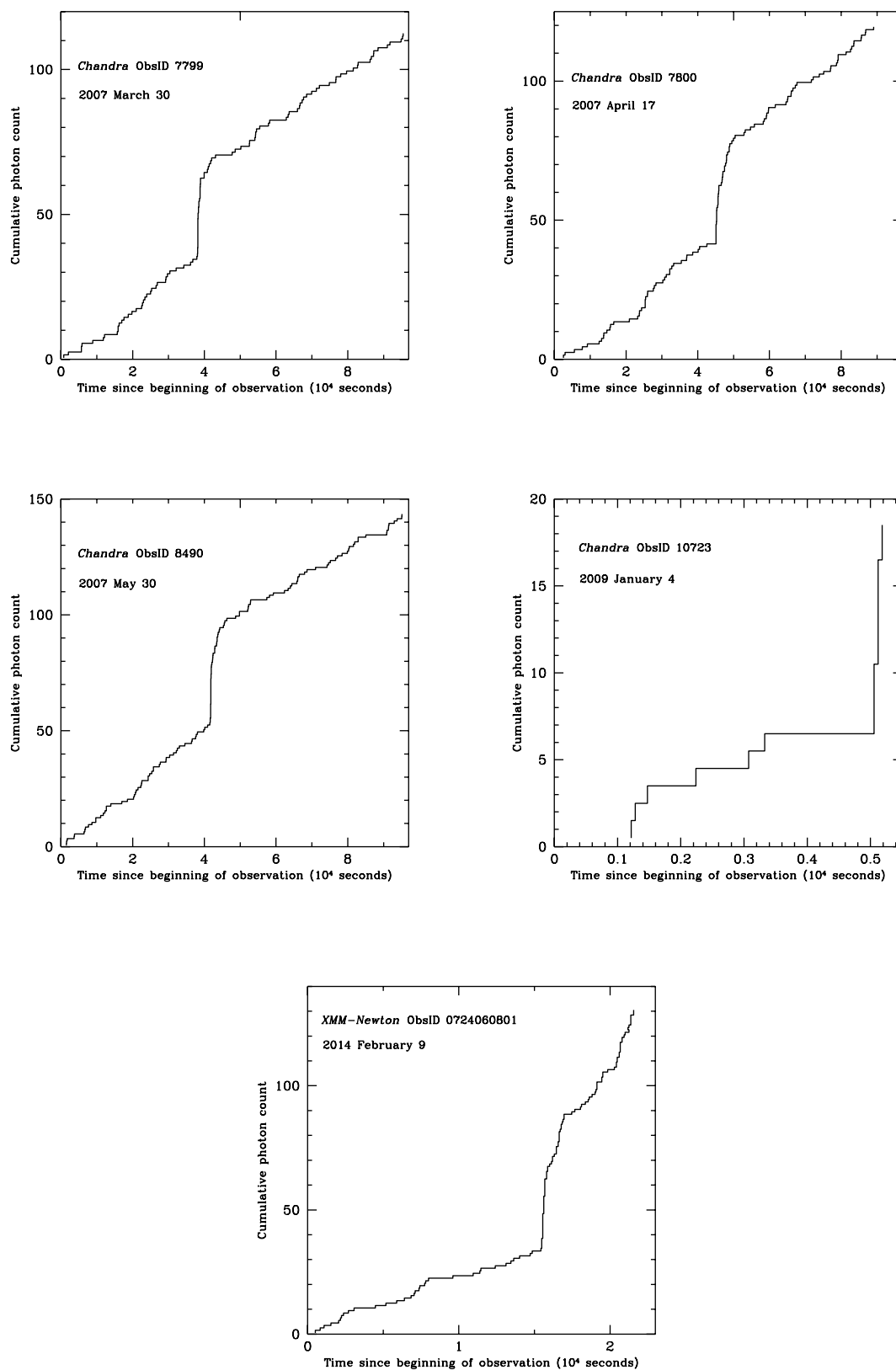
The globular cluster or ultracompact dwarf galaxy hosting Source 2 has a spectroscopically determined metallicity of $Z/H = -0.85 \text{ dex}$ ($Z = 0.14Z_{\odot}$)³². The derived stellar mass¹² of the source is $3.1 \times 10^6 M_{\odot}$. Given its size¹² of 7 pc, and correcting for the luminosity function²⁶ (which predicts that a $4 \times 10^{37} \text{ erg s}^{-1}$ source is ten times more likely to be found in a globular cluster than a $3.2 \times 10^{38} \text{ erg s}^{-1}$ source), we estimate that the globular cluster is expected to have 0.064 X-ray binaries with luminosities of more than $4 \times 10^{37} \text{ erg s}^{-1}$. Therefore, after having found a flaring source, the probability that the persistent emission comes from a different X-ray binary in this cluster is <6.4%. If we conservatively assume that the predicted number of X-ray binaries could be 50% higher (approximately convolving all of the uncertainty sources), then the probability is <9.1%. This might be an overestimate given that ultracompact dwarfs appear to harbour X-ray sources at a lower rate than globular clusters³³.

Even in the most conservative case, the combined probability that both sources arise from sources different from the persistently emitting sources is $<1.5 \times 10^{-4}$.

For both sources, we determined their positions separately during the flare phase and the persistent phase and found no statistical difference within the positional uncertainties. However, this is not highly constraining given the large PSF of Chandra at the off-axis location of the flares.

Code availability. The code used to find X-ray flares is available at <http://pages.astronomy.ua.edu/jairwin/software/>.

25. Dickey, J. M. & Lockman, F. J. H I in the Galaxy. *Annu. Rev. Astron. Astrophys.* **28**, 215–259 (1990).
26. Zhang, Z. *et al.* Luminosity functions of LMXBs in different stellar environments. *Astron. Astrophys.* **533**, A33 (2011).
27. Sivakoff, G. R. *et al.* The low-mass X-ray binary and globular cluster connection in Virgo Cluster early-type galaxies: optical properties. *Astrophys. J.* **660**, 1246–1263 (2007).
28. Kundu, A., Maccarone, T. J. & Zepf, S. E. Probing the formation of low-mass X-ray binaries in globular clusters and the field. *Astrophys. J.* **662**, 525–543 (2007).
29. Harris, W. E. & Harris, G. L. H. The halo stars in NGC 5128. III. An inner halo field and the metallicity distribution. *Astron. J.* **123**, 3108–3123 (2002).
30. Maraston, C. Evolutionary population synthesis: models, analysis of the ingredients and application to high- z galaxies. *Mon. Not. R. Astron. Soc.* **362**, 799–825 (2005).
31. J rdan, A. *et al.* The ACS Virgo Cluster Survey. X. Half-light radii of globular clusters in early-type galaxies: environmental dependencies and a standard ruler for distance estimation. *Astrophys. J.* **634**, 1002–1019 (2005).
32. Beasley, M. A. *et al.* A 2dF spectroscopic study of globular clusters in NGC 5128: probing the formation history of the nearest giant elliptical. *Mon. Not. R. Astron. Soc.* **386**, 1443–1463 (2008).
33. Pandya, V., Mulchaey, J. & Greene, J. E. A comprehensive archival Chandra Search for X-Ray emission from ultracompact dwarf galaxies. *Astrophys. J.* **819**, 162 (2016).



Extended Data Figure 1 | Plots of the cumulative X-ray photon arrival time for the five flares of Source 2 in NGC 5128. The first four flares were observed by Chandra with the fifth flare by XMM-Newton. In ObsID 10723, the first photon of the observation was not received until 1,100 s after the observation began, and the observation ended mid-flare.

Extended Data Table 1 | Summary of the Chandra and XMM-Newton Observations of Sources 1 and 2

Source	Telescope/Detector	ObsID	Observation Date	Exposure (ksec)	Flare?
1	<i>Chandra</i> /ACIS-I	324	1999-12-04	8.5	N
1	<i>Chandra</i> /ACIS-S	323	2000-01-26	53.1	N
1	<i>XMM-Newton</i> /EPIC	0111190101	2000-07-13	27.2	N
1	<i>XMM-Newton</i> /EPIC	0111190501	2000-07-13	6.6	N
1	<i>XMM-Newton</i> /EPIC	0111190201	2000-07-13	66.3	N
1	<i>XMM-Newton</i> /EPIC	0111190701	2001-01-05	64.4	N
1	<i>Chandra</i> /ACIS-I	3926	2003-02-15	75.7	Y
1	<i>Chandra</i> /ACIS-I	4415	2003-02-15	75.3	N
2	<i>Chandra</i> /HRC-I	463	1999-09-10	19.7	N
2	<i>Chandra</i> /HRC-I	1253	1999-09-10	6.9	N
2	<i>Chandra</i> /ACIS-I	316	1999-12-05	36.2	...
2	<i>Chandra</i> /HRC-I	1412	1999-12-21	15.1	N
2	<i>Chandra</i> /HRC-I	806	2000-01-23	65.3	N
2	<i>Chandra</i> /ACIS-I	962	2000-05-17	37.0	N
2	<i>XMM-Newton</i> /EPIC	0093650201	2001-02-02	23.9	N
2	<i>Chandra</i> /ACIS-S/HETG	1600	2001-05-09	47.5	N
2	<i>Chandra</i> /ACIS-S/HETG	1601	2001-05-21	52.2	N
2	<i>XMM-Newton</i> /EPIC	0093650301	2002-02-06	15.3	N
2	<i>Chandra</i> /ACIS-S	2978	2002-09-03	45.2	...
2	<i>Chandra</i> /ACIS-S	3965	2003-09-14	50.2	...
2	<i>Chandra</i> /ACIS-I	7797	2007-03-22	98.2	N
2	<i>Chandra</i> /ACIS-I	7798	2007-03-27	92.0	N
2	<i>Chandra</i> /ACIS-I	7799	2007-03-30	96.0	Y
2	<i>Chandra</i> /ACIS-I	7800	2007-04-17	92.1	Y
2	<i>Chandra</i> /ACIS-I	8489	2007-05-08	95.2	N
2	<i>Chandra</i> /ACIS-I	8490	2007-05-30	95.7	Y
2	<i>Chandra</i> /ACIS-I	10723	2009-01-04	5.2	Y
2	<i>Chandra</i> /ACIS-I	10724	2009-03-07	5.2	N
2	<i>Chandra</i> /HRC-I	10407	2009-04-04	15.2	N
2	<i>Chandra</i> /ACIS-I	10725	2009-04-26	5.0	N
2	<i>Chandra</i> /ACIS-I	10726	2009-06-21	5.2	N
2	<i>Chandra</i> /ACIS-S	10722	2009-09-08	50.0	...
2	<i>Chandra</i> /HRC-I	10408	2009-09-14	15.2	N
2	<i>Chandra</i> /ACIS-I	11846	2010-04-26	4.8	N
2	<i>Chandra</i> /ACIS-I	11847	2010-09-16	5.1	...
2	<i>Chandra</i> /ACIS-I	12155	2010-12-22	5.1	N
2	<i>Chandra</i> /ACIS-I	12156	2011-06-22	5.1	N
2	<i>Chandra</i> /ACIS-I	13303	2012-04-14	5.6	N
2	<i>Chandra</i> /ACIS-I	13304	2012-08-29	5.1	N
2	<i>Chandra</i> /ACIS-I	15294	2013-04-05	5.1	N
2	<i>XMM-Newton</i> /EPIC	0724060501	2013-07-12	12.0	N
2	<i>XMM-Newton</i> /EPIC	0724060601	2013-08-07	12.0	N
2	<i>Chandra</i> /ACIS-I	15295	2013-08-31	5.4	N
2	<i>XMM-Newton</i> /EPIC	0724060701	2014-01-06	26.5	N
2	<i>XMM-Newton</i> /EPIC	0724060801	2014-02-09	23.4	Y
2	<i>Chandra</i> /ACIS-I	16276	2014-04-24	5.1	N
2	<i>Chandra</i> /ACIS-I	16277	2014-09-08	5.4	...
2	<i>Chandra</i> /ACIS-I	17471	2015-03-14	5.4	N
2	<i>Chandra</i> /ACIS-S/LETG	17147	2015-05-13	49.7	N
2	<i>Chandra</i> /ACIS-S/LETG	17657	2015-05-17	50.4	N

'...' indicates that the source of interest did not fall within the field of view of the detector.

Multi-petahertz electronic metrology

M. Garg¹, M. Zhan¹, T. T. Luu¹, H. Lakhotia¹, T. Klostermann¹, A. Guggenmos¹ & E. Goulielmakis¹

The frequency of electric currents associated with charge carriers moving in the electronic bands of solids determines the speed limit of electronics and thereby that of information and signal processing¹. The use of light fields to drive electrons promises access to vastly higher frequencies than conventionally used, as electric currents can be induced and manipulated on timescales faster than that of the quantum dephasing of charge carriers in solids². This forms the basis of terahertz (10¹² hertz) electronics in artificial superlattices², and has enabled light-based switches^{3–5} and sampling of currents extending in frequency up to a few hundred terahertz. Here we demonstrate the extension of electronic metrology to the multi-petahertz (10¹⁵ hertz) frequency range. We use single-cycle intense optical fields (about one volt per ångström) to drive electron motion in the bulk of silicon dioxide, and then probe its dynamics by using attosecond (10^{–18} seconds) streaking^{6,7} to map the time structure of emerging isolated attosecond extreme ultraviolet transients and their optical driver. The data establish a firm link between the emission of the extreme ultraviolet radiation and the light-induced intraband, phase-coherent electric currents that extend in frequency up to about eight petahertz, and enable access to the dynamic nonlinear conductivity of silicon dioxide. Direct probing, confinement and control of the waveform of intraband currents inside solids on attosecond timescales establish a method of realizing multi-petahertz coherent electronics. We expect this technique to enable new ways of exploring the interplay between electron dynamics and the structure of condensed matter on the atomic scale.

Although the inventor of the rectifying diode, Braun, alluded in his Nobel lecture to his unsuccessful attempts to detect light-induced electric currents inside a solid⁸, laser fields are now widely used to control electronic processes in the condensed phase. But whereas optical techniques^{9–11} can induce and track electric currents in solids at frequencies nearing the petahertz (PHz) range, advancing electronic metrology to the multi-petahertz realm calls for the ability to capture the dynamics encoded in radiation emitted at frequencies extending into the extreme ultraviolet (EUV) range and beyond. Here we employ attosecond streaking^{6,7} and photoelectron interferometry¹² to realize multi-petahertz metrology of solids.

Recent spectral-domain studies of laser-driven semiconductors and dielectrics have suggested that the coherent radiation emerging in these interactions (in the visible and ultraviolet¹³, as well as in the extreme ultraviolet¹⁴, EUV) could be directly associated with laser-induced, intraband currents in the bulk, and could therefore serve as a unique macroscopic probe of the microscopic electric currents. Nevertheless, time-resolved studies using mid-infrared and terahertz fields support a different picture; interband polarization, encapsulated in a generalized re-collision model¹⁵ or the interplay^{16,17} between interband and intraband dynamics (Fig. 1a), is essential to describe the nonlinear response of these systems. Theoretical studies^{15,18–21} can now offer valuable insight into these interactions but their conclusions are sensitive to electronic dephasing, which is challenging to account for precisely^{15–17,19–22}. As a result, the question of whether the emerging radiation from the laser-driven solids is linked to the

nonlinear motion of charge carriers in bands (intraband) or to the dipole induced among bands (interband) has been a subject of an escalating debate^{14,18,23,24}. An answer to this question comprises a critical step for extending coherent electronics to the multi-petahertz realm.

To experimentally address this question, we used attosecond streaking to record the temporal profile of EUV transients generated in polycrystalline SiO₂ nanofilms (~120 nm thick) by single-cycle (precisely 1.2-cycle) optical pulses (peak field strength, $F_0 \approx 1.1 \text{ V Å}^{-1}$) produced in a light-field synthesizer²⁵. A streaking spectrogram recorded using our experimental set-up (Fig. 1b; see also Supplementary Information section I) and its numerical reconstruction²⁶ (Supplementary Information section II) are displayed in Fig. 1c and d, respectively. The reconstruction reveals an isolated attosecond EUV pulse, as shown in Fig. 1e, with a duration of $\tau_{\text{EUV}} \approx 470$ as measured at the full-width at half-maximum (FWHM) of its intensity profile; this duration is only slightly longer than the bandwidth-limited value ($\tau_{\text{BL}} \approx 460$ as) and is precisely synchronized to the peak of the driving field. The retrieved spectral phase and spectrum of the attosecond burst are presented in Fig. 1f.

To identify the physical mechanism underlying the nonlinear EUV emission in SiO₂, we performed time–frequency analysis (see Supplementary Information section III) of the retrieved attosecond pulse presented in Fig. 1e, as shown in Fig. 2b. We then compared the results with the nonlinear dipoles obtained through the numerical solution of the semiconductor Bloch equations (SBEs)^{19–21} in SiO₂ including Coulomb interactions among the carriers²¹ and using parameters identical to those in our experiments, such as the retrieved optical field waveform and its strength (Fig. 1e) (see also Supplementary Information section III).

In contrast to earlier SBE modelling of SiO₂ under intense ultrafast fields¹⁴, the incorporation of Coulomb interaction²¹ and the tuning of its strength to reproduce the exciton response of SiO₂ (see Supplementary Information section VI) results in the dominance of the intraband contribution (Fig. 2a, green line) over the interband contribution (Fig. 2a, orange line). Importantly however, the genuine temporal dynamics associated with each of these contributions, earlier identified as macroscopic markers of the physics of the emission^{15,18}, are virtually immune to Coulomb interactions under the conditions of this study (see Supplementary Information section VI). Indeed, interband dynamics gives rise to a positively chirped spectral response (Fig. 2c): that is, EUV photons of different energies are emitted at distinct moments following the excitation of carriers into the bands, their acceleration and their subsequent re-collision¹⁵. By contrast, the contribution of the intraband current yields a virtually concurrent spectral emission (Fig. 2d), which is typical of the nonlinear scattering of a particle in a non-parabolic band, giving rise to a nearly chirp-free EUV dipole synchronized to the peak of the driving field. As the total emission is dominated by the intraband current contribution (Fig. 2a), the chirp of the total EUV dipole (Fig. 2e) is virtually determined by the intraband current.

A comparison between the experimentally traced dynamics (Fig. 2b) and those evaluated for the intraband contribution (Fig. 2d)

¹Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, D-85748 Garching, Germany.

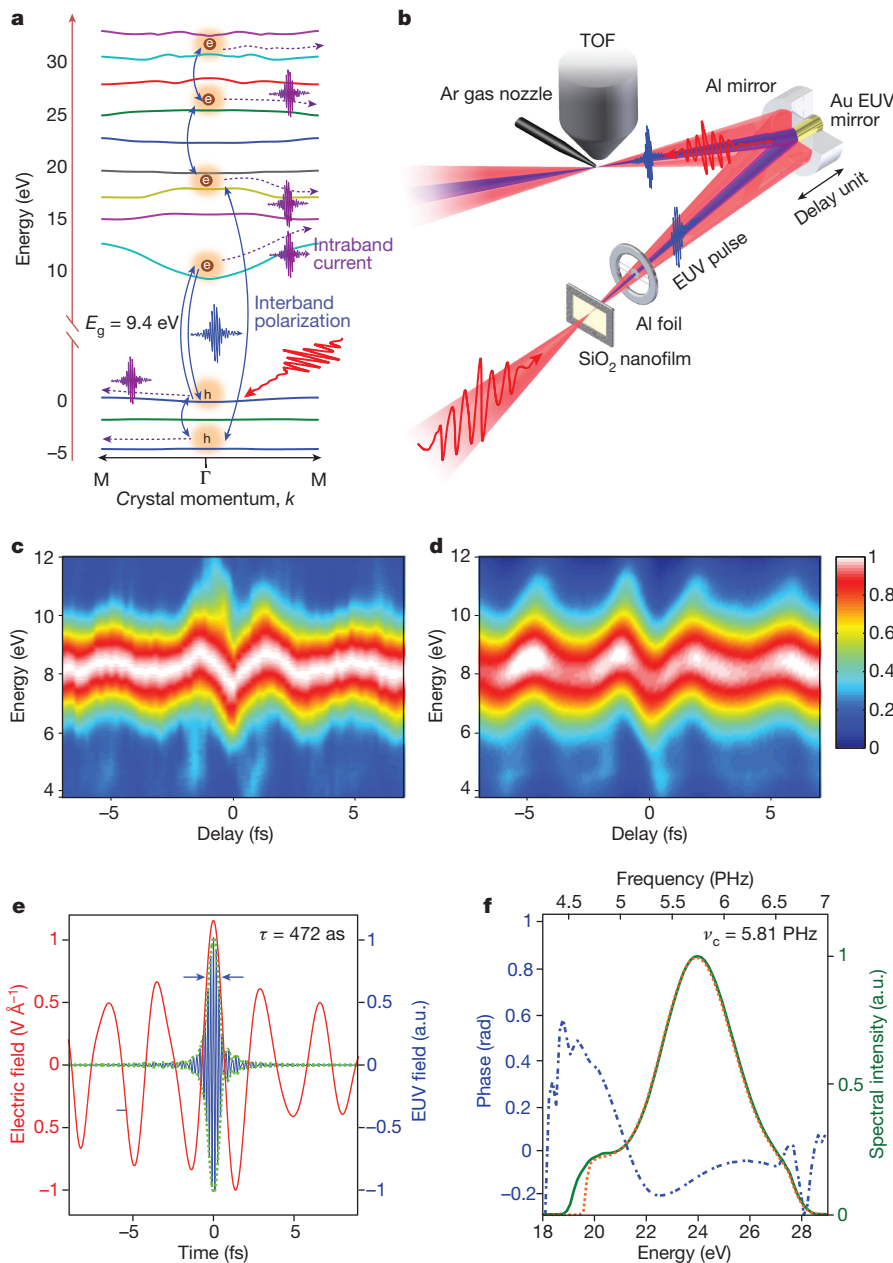


Figure 1 | Attosecond pulse metrology in bulk SiO₂. **a**, Intraband and interband dynamics in optically driven SiO₂ and emission of extreme ultraviolet (EUV) radiation. Intraband currents are induced via the field-driven motion and scattering of electrons (e) and holes (h) along the dispersive band profiles. Dipole coupling between different bands gives rise to an interband contribution to the emitted radiation. **b**, SiO₂ nanofilms (~ 120 nm) are exposed to intense single-cycle optical fields (peak field strength, $F_0 \approx 1.1$ V \AA^{-1}) to generate coherent EUV radiation. A disk-shaped thin Al nanofilm and a concentric two-component, concave mirror-assembly (Al outer, Au inner) allow the spatiotemporal separation of optical and generated EUV pulses as well as their focusing onto an Ar gas jet. A time-of-flight (TOF) spectrometer records photoelectron

spectra generated in Ar as a function of the delay between EUV and the optical pulse. Access to the timing between EUV and optical field at the source is enabled via the absolute delay calibration of the two-component mirror assembly ('delay unit') via a high order autocorrelation of the optical beam (see Supplementary Information section I). **c**, Measured and **d**, reconstructed streaking spectrograms; colour scale represents photoelectron counts in arbitrary units. **e**, Retrieved EUV (blue) and optical (red line) pulse profiles, dashed green line shows the envelope of EUV pulse. **f**, Retrieved spectral phase (blue dashed line) and spectrum (orange dashed line) of the EUV pulse. Green curve indicates the spectrum of the EUV pulse in the absence of the optical probe.

and total polarization (Fig. 2e) reveals a striking agreement, highlighted by the precise synchrony of the peak of the EUV emission with the single intense crest of the optical field and the weak temporal chirp of the generated EUV transients. In contrast, the dynamics of the interband contribution (Fig. 2c) fail to capture the measured attosecond response of the system. A spectrogram (Fig. 2f) simulated using a semiclassical approach (Supplementary Information section IV), in which a pre-excited electron is scattered along the texture of the first conduction

band in the Γ -M direction of a SiO₂ crystal¹⁴, also yields a very good agreement with the experiment.

A detailed comparison with the experimental results (Fig. 2g, black line) of the group delays (emission times) evaluated from the interband contribution (Fig. 2g, orange line), the intraband current (Fig. 2g, green line) and the total dipole (Fig. 2g, orange dashed line) further strengthens the above conclusions, and enables the visualization of the finer details of the dynamic response of the system to the optical

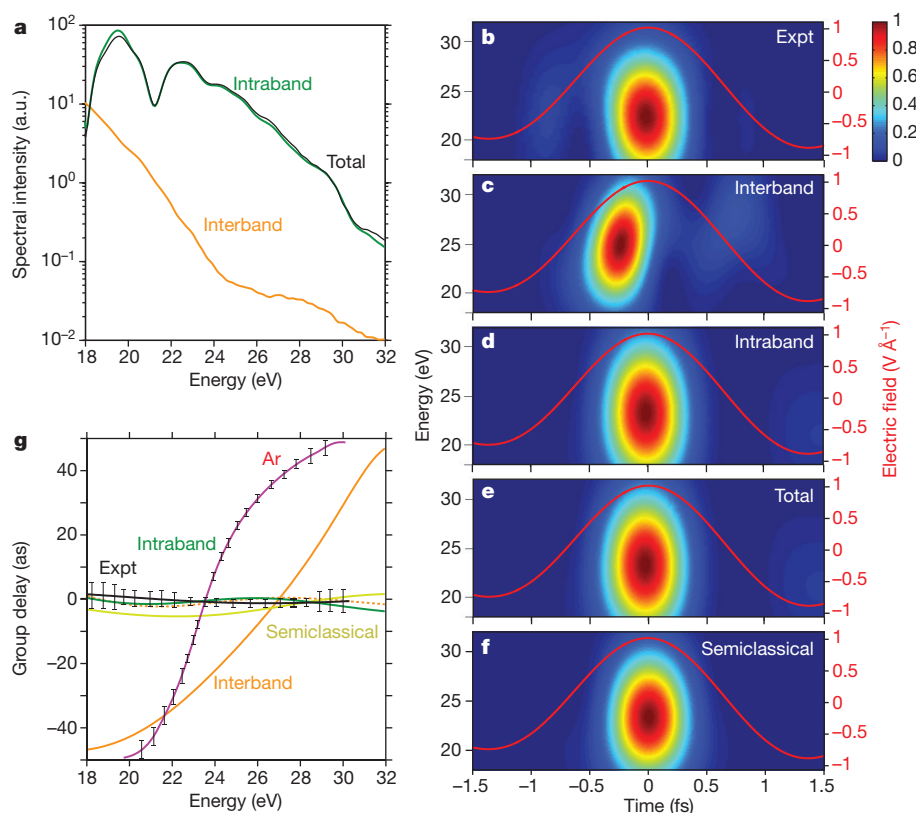


Figure 2 | Interband versus intraband dynamics in SiO₂. **a**, EUV spectra in SiO₂ associated with interband and intraband contributions, and total polarization simulated for the measured optical field shown in Fig. 1e. **b–f**, Time–frequency spectrograms of EUV pulses: **b**, measured in our experiments (Fig. 1e), **c**, interband contribution, **d**, intraband contribution, **e**, total polarization, and **f**, the semiclassical model used in our simulations. Red lines represent the driving electric field (right-

hand y axis). Spectral intensity is shown with the colour scale in arbitrary units. **g**, Retrieved group delays of the EUV emission in SiO₂ ('Expt'; black line) and Ar gas (violet line); standard error in the retrieved group delay is presented with error bars. Shown are group delays of the emission associated with the interband contribution (orange line), the intraband contribution (green line) and total polarization (orange dashed curve), and as modelled using a semiclassical approach (yellow line).

field. Our measurements reveal a weak but statistically significant chirp in the EUV emission that matches the chirp predicted for the intraband current in SBEs and the semiclassical model. Experiments under identical conditions performed in Ar (see also Supplementary Information section I) reproduced earlier recognized chirp-dynamics of gas-phase EUV emission^{27,28}, as shown in Fig. 2g (violet curve), which is in very good agreement with the predictions of semiclassical theory (Supplementary Fig. 10) and serve here as a benchmark of the temporal resolution of our apparatus. Additional measurements in crystalline SiO₂ (see Supplementary Information section VII) as well as measurements performed at different field strengths of the optical driver further substantiate the above conclusions (see Supplementary Information section V).

In a second set of experiments, we probe the nonlinearity of the emission process directly in the time domain, by using attosecond streaking to trace the temporal structure of EUV bursts generated by single-cycle pulses whose carrier envelope phase (φ_{CE}) is adjusted to produce near-cosine and near-sine waveforms (white lines in Fig. 3a, b). The contrast in field strengths between the two most intense half-cycles in each of these waveforms was 1.6 to 1 (Fig. 3a) and 1.16 to 1 (Fig. 3b), respectively. For both the above driving waveforms the experiments reveal the emergence of an isolated burst of radiation (Fig. 3c, d), which is precisely synchronized to their most intense field crest. This feature is well reproduced in our simulations by intraband current (Fig. 3e, f, green lines) and total polarization (Fig. 3g, h), and it is compatible with the findings of previous time-integrated studies¹⁴ in which an $\sim F^{18}$ scaling of the EUV emission versus the electric field of the driving pulse (F) was identified. Based on that nonlinearity, a contrast of several orders of magnitude between main and satellite burst is predicted for

both φ_{CE} values studied here, in agreement with our experiments. The onset and disappearance of a pulse pedestal (Fig. 3c, d, black arrows) is compatible with a φ_{CE} -induced spectral shaping of the EUV emission and is well reproduced by our theoretical modelling via intraband dynamics (Fig. 3e, f; see also Supplementary Information section VIII). This dynamic variation of properties of the EUV pulse via the φ_{CE} of the driver demonstrates the capability of controlling the frequency and the time structure of the laser-induced electric current in a solid. Dynamics pertaining to the interband contribution in our model suggest the generation of a single EUV burst for the near-cosine pulse (Fig. 3e) but also the formation of a sizable satellite burst for a near-sine pulse (Fig. 3f); both are generated at a time offset with respect to the peaks of the driving fields, as presented in Fig. 3e, f. These dynamics bear a similarity to those revealed in φ_{CE} studies of EUV pulse generation in gases at the single-cycle limit²⁸, but do not match the experimental findings of the present study.

Attosecond streaking has thus far provided direct access to the envelope and frequency sweep (chirp) of the multi-petahertz electric currents, but the establishment of multi-petahertz electronics requires access to, and confirmation of, the phase coherence of these currents: that is, the reproducibility of their instantaneous waveform from pulse to pulse and immunity of this waveform to φ_{CE} fluctuations of the optical driver. So far, access to the phase coherence of EUV pulse trains has been demonstrated in gases²⁹, but in solids, and at the isolated attosecond pulse limit, phase coherence has remained unexplored. To study the waveform reproducibility of our EUV pulses, we have employed an earlier proposed methodology¹² in which φ_{CE} dynamics can be accessed via the interference in photoelectron spectra, generated in a gas jet (here Ar) via the direct photoionization by the EUV pulse

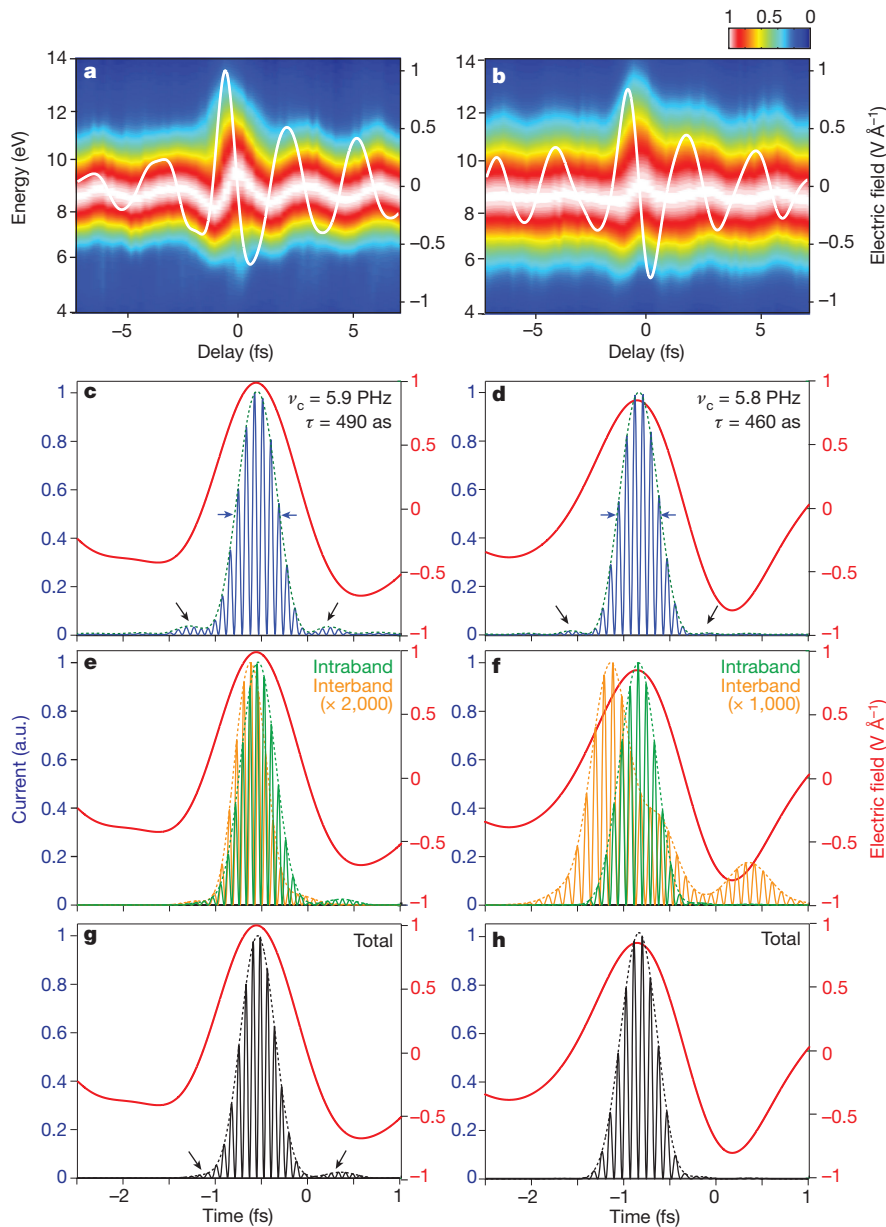


Figure 3 | Control of multi-petahertz currents in SiO₂. **a, b,** Attosecond streaking spectrograms recorded for two φ_{CE} settings of the single-cycle driving pulse (white lines; right-hand y axis), differing by $\sim\pi/2$ rad; colour scale shows photoelectron counts in arbitrary units. **c, d,** EUV transients (instantaneous intensity) retrieved from the spectrograms shown in **a** and **b**, respectively. Blue and red lines represent the intensity profile and field waveform of EUV transients and optical driver respectively, green dashed curve shows the envelope of EUV transients. ν_c and τ are carrier frequency and pulse duration of the EUV transients respectively. **e, f,** Instantaneous intensity profiles of EUV pulses predicted for the intraband contribution (green curve) in our simulation for experimentally recorded optical waveforms (red lines). Interband contributions (orange line) are scaled in **e, f**, to ease comparison with intraband contributions (green line). **g, h,** Total polarization dynamics (black lines) simulated for the two driving waveforms (red lines). Black arrows highlight the temporal features in EUV transients in **c** and **g**.

and through above-threshold ionization (ATI) by the optical field (see Supplementary Information, section X). Figure 4a shows such interference fringes recorded over a time period of 15 min. The φ_{CE} stability evaluated from the interference pattern (Fig. 4b) was better than $\sim\pi/10$ rad, as displayed in Fig. 4c, implying that the EUV waveform and thus that of the intraband current are reproducible with an accuracy better than one-twentieth of their carrier period. Detailed control of the φ_{CE} phase of the EUV pulses (see Supplementary Information section X) further verifies the accuracy of our approach. For a nonlinear medium with a length less than the driving pulse wavelength, the field profiles of the EUV pulse and that of the intraband current are related as: $E_{EUV}(t) \propto J(t)$ (see Supplementary Information section XI). As a result, the EUV fields displayed in Fig. 1e, Fig. 3a, b and Supplementary Fig. 17 comprise the first demonstration of the use of optical pulses to generate, measure, confine (at the attosecond level) and control waveform-reproducible multi-petahertz currents in solids.

To gain quantitative insight into the electronic properties of SiO₂ in the multi-petahertz regime, we measured the photon yield of the emitted EUV light, which in turn allowed us to evaluate the amplitude of $E_{EUV}(t)$ in the bulk of SiO₂ (see Supplementary Information section XI) as well as the nonlinear current density in our sample as shown

in Fig. 4d. The current density reaches values of $\sim 10^{11}$ A m⁻², which is approximately an order of magnitude higher than earlier efforts in sub-petahertz ranges³. The dynamic nonlinear conductivity³⁰ $\sigma(t)$ shown in Fig. 4e, which is evaluated from the precise knowledge of driving optical field $F(t)$, current density profile $j(t)$ and their relative timing as $\sigma(t) = j(t)/F(t)$, builds up within approximately ~ 0.7 fs and switches periodically at the frequency of the generated EUV field: that is, it turns on and off within approximately ~ 30 as (Fig. 4e inset). Negative values of the dynamic conductivity imply that electrons move in opposite directions with respect to the applied field force every second half cycle of their oscillation (Fig. 4e); such a regime was earlier recognized in intraband coherent electron motions in semiconductor superlattices², and is here extended into the multi-petahertz range.

Identifying, measuring in real-time and controlling multi-petahertz intraband electric currents in solids, and similarly understanding the dynamic electronic properties of solids in the multi-petahertz regime, opens up new prospects of study and applications at the interface of photonics and electronics. Owing to the high sensitivity of these currents to the atomic-scale structure of materials (which in turn dictates the details of the dispersion profiles of the electronic bands¹⁴), laser-induced intraband currents and their probing via the emitted

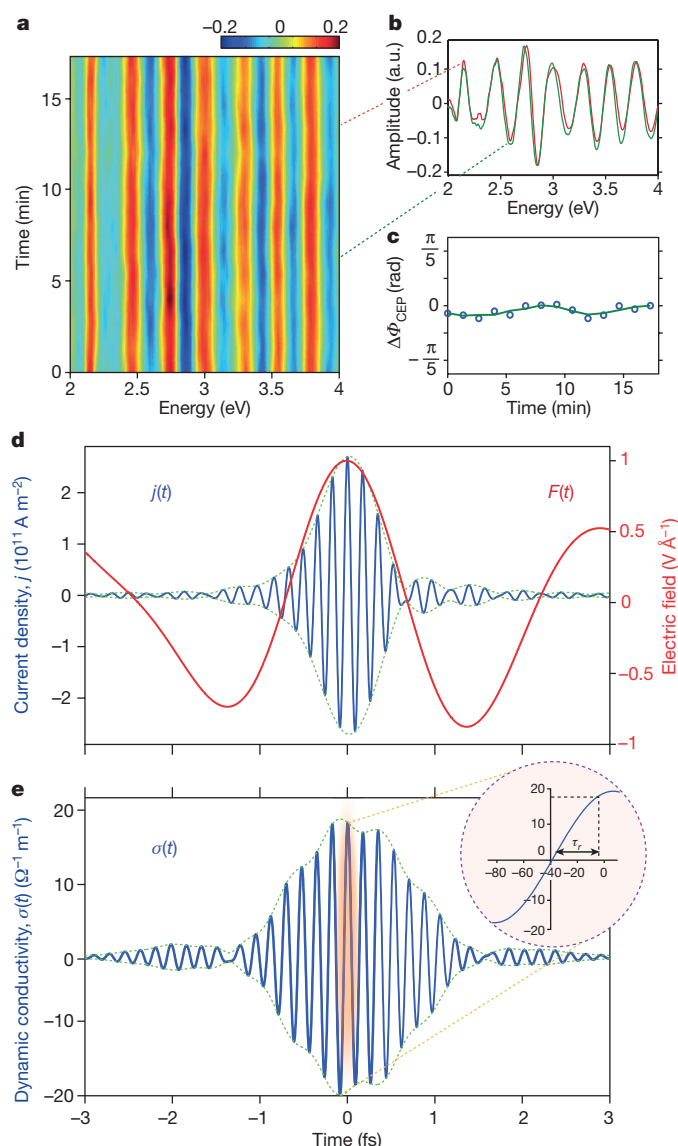


Figure 4 | Phase coherence of multi-petahertz currents and the dynamic conductivity of SiO₂. **a**, Spectral interference between ATI and EUV photoelectrons in Ar, recorded for a time period of ~ 15 min, colour scale represents spectral intensity in arbitrary units. **b**, Interference pattern (red and green lines) at two representative instances of the measurement in **a**. **c**, Evaluated φ_{CEP} (blue circles) variation of the EUV pulse (and the corresponding intraband current), green line is to guide the eye. **d**, Nonlinear current density, $j(t)$ (blue curve). Red curve shows the driving optical field, $F(t)$, green dashed curve shows the envelope of $j(t)$. **e**, Induced nonlinear dynamic conductivity $\sigma(t)$ in SiO₂ (blue curve). Inset highlights the rapid switching of the conductivity within $\tau_r \approx 30$ as. Vertical axis, as main panel; horizontal axis, time in as.

EUV transients may soon enable direct probing of the periodic potentials of solids.

Received 26 February; accepted 30 August 2016.

- Caulfield, H. J. & Dolev, S. Why future supercomputing requires optics. *Nat. Photon.* **4**, 261–263 (2010).
- Leo, K. *High-Field Transport in Semiconductor Superlattices* (Springer, 2003).
- Schiffrin, A. *et al.* Optical-field-induced current in dielectrics. *Nature* **493**, 70–74 (2012).

- Krüger, M., Schenk, M. & Hommelhoff, P. Attosecond control of electrons emitted from a nanoscale metal tip. *Nature* **475**, 78–81 (2011).
- Somma, C., Reimann, K., Flytzanis, C., Elsaesser, T. & Woerner, M. High-field terahertz bulk photovoltaic effect in lithium niobate. *Phys. Rev. Lett.* **112**, 146602 (2014).
- Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* **88**, 173903 (2002).
- Goulielmakis, E. *et al.* Direct measurement of light waves. *Science* **305**, 1267–1269 (2004).
- Braun, F. Electrical oscillations and wireless telegraphy. In *Nobel Lectures, Physics 1901–1921* (Elsevier, 1967).
- Gaal, P. *et al.* Internal motions of a quasiparticle governing its ultrafast nonlinear response. *Nature* **450**, 1210–1213 (2007).
- Huber, R. *et al.* How many-particle interactions develop after ultrafast excitation of an electron-hole plasma. *Nature* **414**, 286–289 (2001).
- Gudde, J., Rohleder, M., Meier, T., Koch, S. W. & Hofer, U. Time-resolved investigation of coherently controlled electric currents at a metal surface. *Science* **318**, 1287–1291 (2007).
- Liu, C. D. *et al.* Carrier-envelope phase effects of a single attosecond pulse in two-color photoionization. *Phys. Rev. Lett.* **111**, 123901 (2013).
- Ghimire, S. *et al.* Observation of high-order harmonic generation in a bulk crystal. *Nat. Phys.* **7**, 138–141 (2011).
- Luu, T. T. *et al.* Extreme ultraviolet high-harmonic spectroscopy of solids. *Nature* **521**, 498–502 (2015).
- Vampa, G. *et al.* Linking high harmonics from gases and solids. *Nature* **522**, 462–464 (2015).
- Hohenleutner, M. *et al.* Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).
- Schubert, O. *et al.* Sub-cycle control of terahertz high-harmonic generation by dynamical Bloch oscillations. *Nat. Photon.* **8**, 119–123 (2014).
- Wu, M. X., Ghimire, S., Reis, D. A., Schafer, K. J. & Gaarde, M. B. High-harmonic generation from Bloch electrons in solids. *Phys. Rev. A* **91**, 043839 (2015).
- Kira, M. & Koch, S. W. *Semiconductor Quantum Optics* (Cambridge Univ. Press, 2012).
- Golde, D., Meier, T. & Koch, S. W. High harmonics generated in semiconductor nanostructures by the coupled dynamics of optical inter- and intraband excitations. *Phys. Rev. B* **77**, 075330 (2008).
- Haug, H. & Koch, S. W. *Quantum Theory of the Optical and Electronic Properties of Semiconductors* 5th edn (World Scientific, 2009).
- Schultze, M. *et al.* Attosecond band-gap dynamics in silicon. *Science* **346**, 1348–1352 (2014).
- McDonald, C. R., Vampa, G., Corkum, P. B. & Brabec, T. Interband Bloch oscillation mechanism for high-harmonic generation in semiconductor crystals. *Phys. Rev. A* **92**, 033845 (2015).
- Tamaya, T., Ishikawa, A., Ogawa, T. & Tanaka, K. Diabatic mechanisms of higher-order harmonic generation in solid-state materials under high-intensity electric fields. *Phys. Rev. Lett.* **116**, 016601 (2016).
- Hassan, M. T. *et al.* Optical attosecond pulses and tracking the nonlinear response of bound electrons. *Nature* **530**, 66–70 (2016).
- Mairesse, Y. & Quere, F. Frequency-resolved optical gating for complete reconstruction of attosecond bursts. *Phys. Rev. A* **71**, 011401(R) (2005).
- Corkum, P. B. Plasma perspective on strong-field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Goulielmakis, E. *et al.* Single-cycle nonlinear optics. *Science* **320**, 1614–1617 (2008).
- Benko, C. *et al.* Extreme ultraviolet radiation with coherence time greater than 1 s. *Nat. Photon.* **8**, 530–536 (2014).
- Mics, Z. *et al.* Thermodynamic picture of ultrafast charge transport in graphene. *Nat. Commun.* **6**, 7655 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a European Research Council grant (Attoelectronics-258501), the Deutsche Forschungsgemeinschaft Cluster of Excellence, Munich Centre for Advanced Photonics, the Max Planck Society and the European Research Training Network MEDEA.

Author Contributions M.G. and M.Z. conducted the experiments; E.G. planned the experiments and supervised the project; M.G., H.L., T.K., T.T.L. and A.G. conducted the simulations; and M.G. and E.G. interpreted the experimental data and contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (elgo@mpq.mpg.de).

Reviewer Information Nature thanks M. Chini, U. Höfer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Real-space investigation of energy transfer in heterogeneous molecular dimers

Hiroshi Imada¹, Kuniyuki Miwa¹, Miyabi Imai-Imada^{1,2}, Shota Kawahara^{1,2}, Kensuke Kimura^{1,2} & Yousoo Kim¹

Given its central role in photosynthesis^{1–4} and artificial energy-harvesting devices^{5–7}, energy transfer has been widely studied using optical spectroscopy to monitor excitation dynamics and probe the molecular-level control of energy transfer between coupled molecules^{2–4}. However, the spatial resolution of conventional optical spectroscopy is limited to a few hundred nanometres and thus cannot reveal the nanoscale spatial features associated with such processes. In contrast, scanning tunnelling luminescence spectroscopy^{8–19} has revealed the energy dynamics associated with phenomena ranging from single-molecule electroluminescence^{11,12,14,17,19}, absorption of localized plasmons¹⁹ and quantum interference effects^{19–21} to energy delocalization¹⁷ and intervalley electron scattering¹⁵ with submolecular spatial resolution in real space. Here we apply this technique to individual molecular dimers that comprise a magnesium phthalocyanine and a free-base phthalocyanine (MgPc and H₂Pc) and find that locally exciting MgPc with the tunnelling current of the scanning tunnelling microscope generates a luminescence signal from a nearby H₂Pc molecule as a result of resonance energy transfer from the former to the latter. A reciprocating resonance energy transfer is observed when exciting the second singlet state (S₂) of H₂Pc, which results in energy transfer to the first singlet state (S₁) of MgPc and final funnelling to the S₁ state of H₂Pc. We also show that tautomerization²² of H₂Pc changes the energy transfer characteristics within the dimer system, which essentially makes H₂Pc a single-molecule energy transfer valve device that manifests itself by blinking resonance energy transfer behaviour.

A scanning tunnelling microscope (STM) combined with an optical system is an ideal platform for local investigation of energy dynamics, which provides analysis capabilities of geometric and electronic structures as well as excitation and relaxation characteristics⁹. Figure 1a sketches our experimental set-up for investigating energy transfer between two different molecules (donor and acceptor): the atomically confined tunnelling current of the STM excites only the donor molecule, and the photons emitted from the coupled molecular system are then detected. MgPc and H₂Pc fluoresce at separate energies²³ and were selected as the sample molecules (Fig. 1b). Figure 1c shows the topographic STM image of the sample, where MgPc, H₂Pc and CO molecules are co-deposited on an ultrathin NaCl(100) film grown on Ag(111). The use of an NaCl film as the substrate decouples the molecules from the metallic substrate and enables the optical investigation of single molecules using the STM^{11,17,19}.

The adsorption structures of MgPc and H₂Pc on the NaCl film were determined with atomic precision using a CO-terminated STM tip (Fig. 1d)^{24,25}; the results are summarized with a detailed analysis based on density functional theory (DFT) and reported elsewhere²⁶. Briefly, the centre of H₂Pc adsorbs on a sodium ion with the molecular axes aligned in the [010] or [001] directions, whereas the centre of MgPc adsorbs on a chlorine ion with the molecular axes tilted $\pm 38^\circ$ from the [010] or [001] directions. The peculiar 16-lobe appearance of an isolated MgPc molecule is due to a rapid shuttling motion between the

two equivalently stable adsorption angles ($+38^\circ \leftrightarrow -38^\circ$; ref. 26). The configuration of a MgPc–H₂Pc dimer can be specified with a vector connecting the molecular centres of H₂Pc and MgPc, where the unit vectors α and β are between the nearest neighbour chlorines (Fig. 1d). Two different MgPc–H₂Pc dimers with (5.5, 2.5) and (3.5, 2.5) configurations were investigated. Notably, the shuttling motion of MgPc is stopped in the (3.5, 2.5) configuration because the two angles ($\pm 38^\circ$) are no longer equivalent (see Extended Data Fig. 1), resulting in the eight-lobe appearance that is typical for many phthalocyanines²⁷.

Figure 1e shows the scanning tunnelling luminescence (STL) spectra obtained with the STM tip placed on the MgPc in the two MgPc–H₂Pc dimers and those of isolated MgPc and H₂Pc molecules as references. The MgPc molecule shows a fluorescence peak from the S₁ state (the so-called Q state) at 1.89 eV and its vibronic satellites on the lower-energy side. Because the presence of two hydrogen atoms at the centre of H₂Pc lowers the molecular symmetry, the Q state—which is doubly degenerate in many fourfold symmetric phthalocyanines—splits into two excited states, Q_x and Q_y^{19,28}. The H₂Pc molecule therefore shows a sharp and intense fluorescence peak from the S₁ state (Q_x) at 1.81 eV and a weak fluorescence from S₂ (Q_y) at 1.92 eV (ref. 19). The small peaks at approximately 1.60–1.75 eV are the vibronic satellites. The Q_x and Q_y states are polarized in the x and y directions (Fig. 1b)^{19,28}, respectively.

In the (5.5, 2.5) dimer with an intermolecular distance of 2.4 nm, the sharp fluorescence peak of H₂Pc at 1.81 eV was observed while locally exciting the MgPc. The H₂Pc Q_x fluorescence peak became pronounced in the (3.5, 2.5) configuration with a shorter intermolecular distance of 1.7 nm. The intensity increase in H₂Pc fluorescence accompanied a decrease in MgPc Q fluorescence, clearly indicating energy transfer from MgPc to H₂Pc. This energy transfer is not very sensitive to the tip position, and can be observed when placing the tip at any point over the MgPc molecule (Extended Data Fig. 2). Another important observation is that H₂Pc Q_y fluorescence was not detected even in the (3.5, 2.5) configuration.

To elucidate the mechanism of the energy transfer in the MgPc–H₂Pc dimer, the electronic structure was investigated using scanning tunnelling spectroscopy (STS). Figure 2a shows the dI_t/dV (where I_t is the tunnelling current and V is the sample bias voltage) spectra of the MgPc and H₂Pc molecules. They show two peaks, one for positive sample voltages and the other for negative voltages, corresponding to resonant tunnelling through the lowest unoccupied molecular orbital (LUMO) and highest occupied molecular orbital (HOMO), respectively²⁹. They have similar energy gaps between the resonance channels; however, the dI_t/dV peaks of MgPc are located 300 mV higher than those of H₂Pc. This energy level configuration is almost intact on formation of the dimer. Figure 2b shows the spatial variation of dI_t/dV signals across the MgPc–H₂Pc heterojunction. This clearly indicates that their molecular orbitals are not strongly hybridized with each other, consistent with the results of DFT analysis (Extended Data Fig. 3).

The threshold voltage to induce the MgPc fluorescence was observed at -1.95 V (Extended Data Fig. 4), which matches with the upper edge

¹Surface and Interface Science Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. ²Department of Advanced Materials Science, Graduate School of Frontier Science, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8651, Japan.

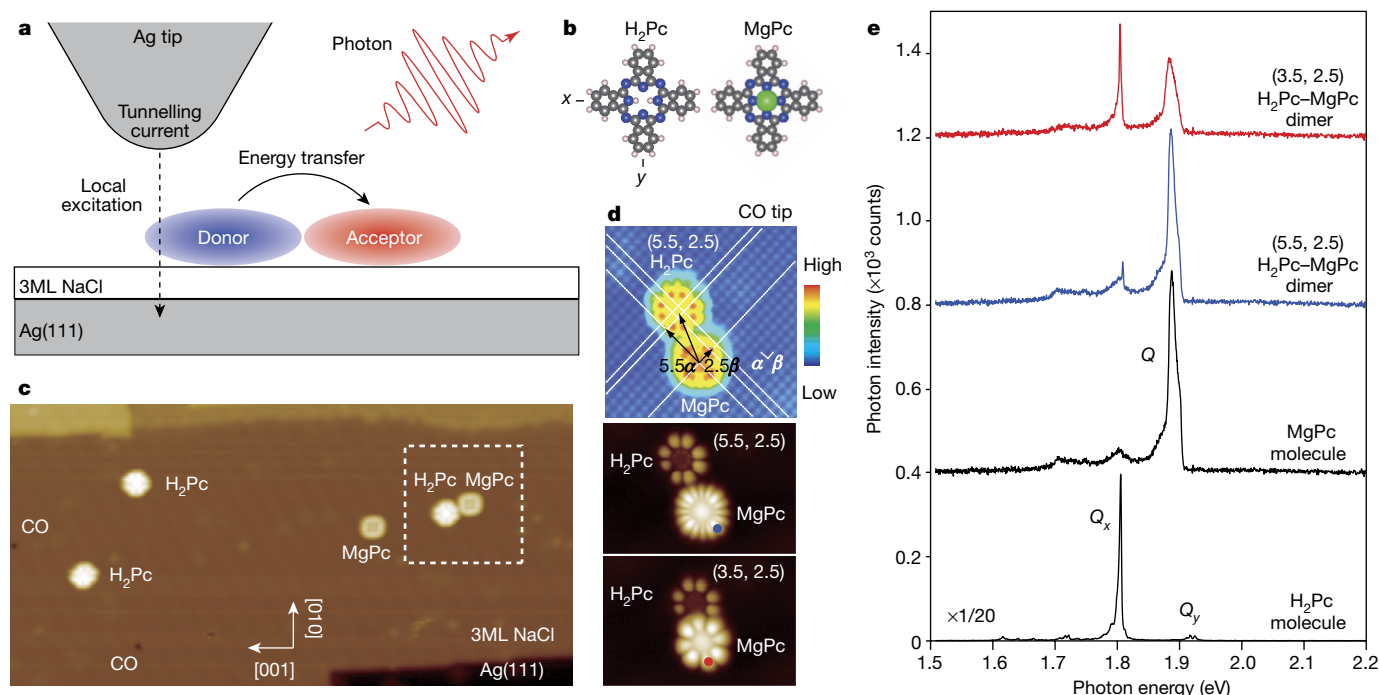


Figure 1 | Excitation of molecular fluorescence through intermolecular energy transfer. **a**, The design of the experiment to investigate energy transfer between two different molecules. **b**, Structures of H₂Pc and MgPc (grey, C; blue, N; white, H; and green, Mg). The *x* axis of H₂Pc is parallel to the N–H–N bond²⁸. **c**, An STM image of the sample molecules on a three-monolayer (ML)-thick NaCl(100) island on Ag(111) (sample bias voltage $V = 1$ V, tunnelling current $I_t = 5$ pA, 65×32.5 nm²). **d**, STM images of the area surrounded by the dashed line in **c** (top: $V = -2.5$ V, $I_t = 2$ pA, with a

CO-terminated tip, middle and bottom: $V = -2.3$ V, $I_t = 5$ pA, with a metal tip). The vector in each STM image is the specification of the dimer configuration. For the middle to bottom images the H₂Pc was moved to change the dimer configuration. **e**, STL spectra of the H₂Pc molecule, the MgPc molecule, the (5.5, 2.5) dimer and the (3.5, 2.5) dimer. The red and blue curves were measured at the red and blue points in **d**. The measurement conditions for the H₂Pc molecule were $V = -2.3$ V, $I_t = 30$ pA and exposure time $t = 1$ min, and the conditions for the others were $V = -2.1$ V, $I_t = 30$ pA and $t = 1$ min.

of the tunnelling channel through the HOMO of MgPc. Therefore, the fluorescence of MgPc is triggered by hole injection into the HOMO of MgPc¹⁷. Figure 2c shows the carrier and energy dynamics in the

MgPc–H₂Pc dimer induced by the injection of a hole into the HOMO of MgPc. The injected hole remains there, because the hole transfer to H₂Pc is blocked by the energy barrier at the MgPc–H₂Pc junction.

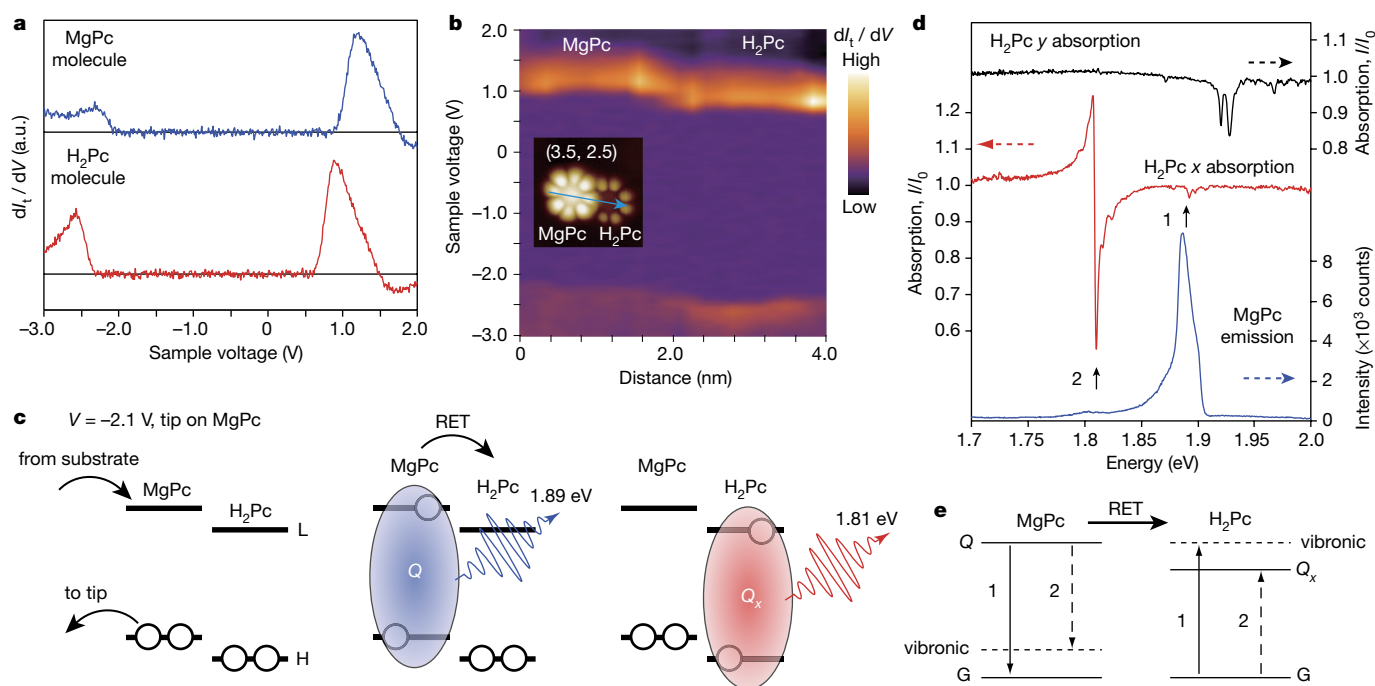


Figure 2 | Energy-level alignment and the mechanism of energy transfer. **a**, dI_t/dV spectra of H₂Pc and MgPc molecules. **b**, Twenty-five dI_t/dV spectra were measured across the (3.5, 2.5) MgPc–H₂Pc junction along the blue arrow in the inset (set point: $V = -2.6$ V, $I_t = 5$ pA). **c**, Schematics illustrating the dynamic process. It was assumed that the

STM tip was on the MgPc with $V = -2.1$ V. **d**, Spectral overlap of the MgPc emission (blue curve) and the H₂Pc absorptions (red and black). The single-molecule absorption spectra of the H₂Pc molecule were obtained using the technique developed in ref. 19. **e**, Schematic of the two main RET pathways. G, ground state; H, HUMO; L, LUMO.

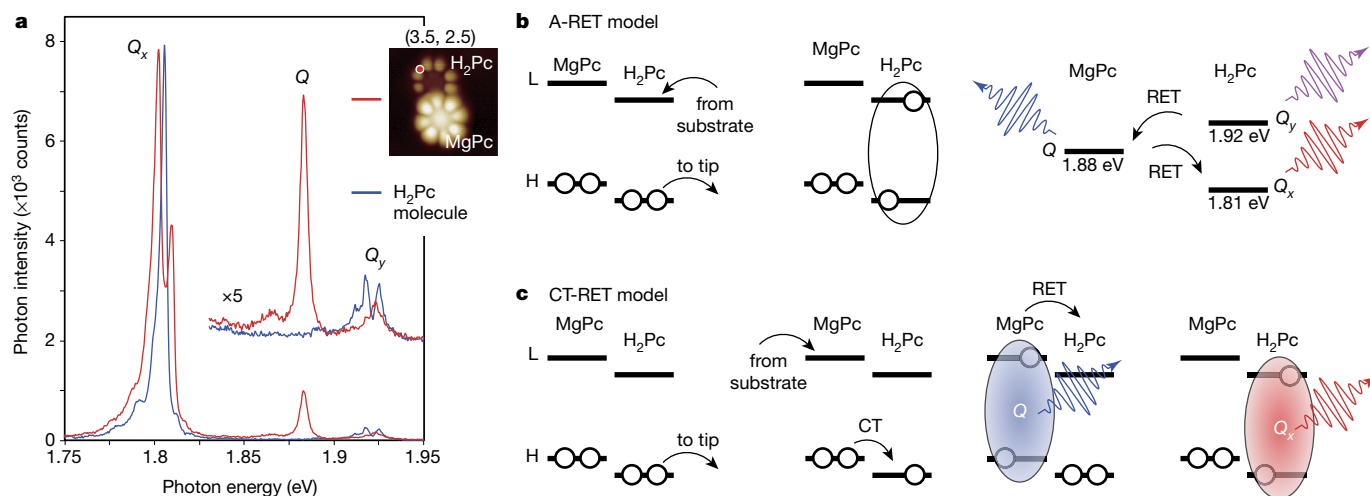


Figure 3 | Backward energy transfer from H₂Pc to MgPc. **a**, STS spectra measured on an H₂Pc molecule (blue) and H₂Pc in a (3,5,2,5) MgPc-H₂Pc dimer (red: measured at the red point in the inset) with $V = -2.3$ V, $I_t = 30$ pA, $t = 1$ min. **b**, **c**, The A-RET (**b**) and CT-RET (**c**) models used to explain the observations. H, HOMO; L, LUMO.

The hole in the HOMO of MgPc is most probably filled by electron transfer from the metal substrate. However, there is a non-negligible probability of electron supply to the LUMO of MgPc from the metal substrate, leading to exciton formation in MgPc. This is rationalized by the reduction of the electron injection barrier into the LUMO, which is induced by hole injection into the HOMO¹⁶ and the potential drop in the NaCl film¹⁷. Once the Q state is formed in MgPc, energy transfer by charge transfer is again prohibited by the energy barrier. The Q state of MgPc (1.89 eV) is higher in energy than the Q_x state (1.81 eV) of H₂Pc, but lower than the Q_y state (1.92 eV) of H₂Pc; therefore, the only possible mechanism for energy transfer is resonance energy transfer (RET) from the Q state of MgPc to the Q_x state of H₂Pc.

The spectral overlap of the fluorescence of MgPc and the absorption of H₂Pc was examined to reveal RET pathways (Fig. 2d). Single-molecule absorption spectra of H₂Pc reveal the electronic transitions polarized in the *x* direction (*x* absorption) and those polarized in the *y* direction (*y* absorption)¹⁹. In addition to the strong features at 1.81 eV in *x* and at 1.92 eV in *y*, vibronic satellites of each transition are clearly seen in the higher energy sides. There is evidence of two main RET pathways: from Q to the vibrationally excited state of Q_x, and from Q to the vibrational ground state of Q_x, which accompanies a vibronic excitation in MgPc (indicated as two upward arrows in Fig. 2d and illustrated in Fig. 2e). The insensitivity of the RET to the tip position (Extended Data Fig. 2)

suggests that the intermolecular dipole-dipole interaction that dominates the RET process is not greatly influenced by the plasmonic field and is determined mostly by the relative configuration of the two transition dipole moments of the energy-donating and -accepting excited states.

Next, the backwards energy transfer from H₂Pc to MgPc was investigated (Fig. 3). The STS spectrum measured on the H₂Pc in the MgPc-H₂Pc dimer shows distinct features from that of H₂Pc molecule: (1) the Q_x peak was split into two peaks with an energy separation of 6 meV; (2) the total number of photons from the Q_x state (integrated in the energy range 1.78–1.82 eV) increased by a factor of 1.6; (3) Q fluorescence appeared at 1.88 eV; and (4) the intensity of the Q_y fluorescence decreased, but was not completely quenched. The peak splitting of the Q_x fluorescence (1) can be attributed to the slightly different energy levels in the two most stable MgPc-H₂Pc dimers (Extended Data Fig. 3) and current-induced tautomerization²² between them. Surprisingly, the observations (2–4) indicate a successive energy transfer Q_y → Q → Q_x.

Two plausible models are proposed to explain the observations. The first model is an avalanche RET (A-RET) model. The exciton (either Q_x or Q_y) is created in the H₂Pc by hole injection into the HOMO of H₂Pc (Extended Data Fig. 5) and electron transfer to the LUMO or LUMO+1 of H₂Pc from the substrate. If a Q_x exciton is created, it remains in the state because other excited states are higher in energy. If a Q_y exciton is created, it undergoes energy transfer to the Q state of MgPc by RET,

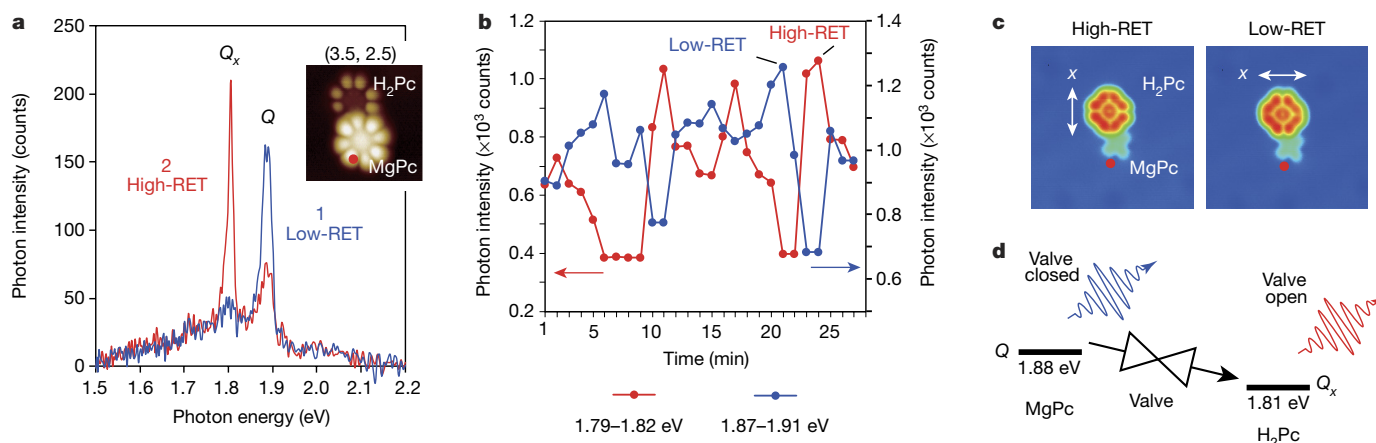


Figure 4 | Blinking of RET. **a**, Twenty-seven consecutive STS spectra were measured at the fixed tip position on MgPc in the (3,5,2,5) dimer ($V = -2.1$ V, $I_t = 5$ pA, $t = 1$ min, at the red dot in the inset). The blue and red curves were measured at 21 min and 24 min, respectively. **b**, Photon intensities integrated over 1.79–1.82 eV (red) and 1.87–1.91 eV

(blue) are plotted as a function of time. **c**, Two STM images ($V = 0.7$ V) corresponding to the molecular configurations of high- and low-RET probabilities, respectively. **d**, A schematic diagram of the valve operation (open/closed) for RET achieved with H₂Pc.

followed by another RET to the Q_x state of H_2Pc . The A-RET model explains all of the main observations: the reduced but non-zero Q_y fluorescence, the emergence of the Q fluorescence and the increase in the Q_x fluorescence. The second model (the charge-transfer (CT)-RET model) involves hole transfer from the HOMO of H_2Pc to the HOMO of $MgPc$ and exciton creation in $MgPc$, followed by RET to the Q_x state of H_2Pc . The CT-RET model can explain the emergence of the Q fluorescence; however, the non-zero Q_y fluorescence is not consistent with the CT-RET model alone. In conclusion, the A-RET model is the main process for the back-and-forth energy transfer; however, the CT-RET process cannot be completely excluded. The dominant contribution of the A-RET model is also supported by the fact that the Q fluorescence of $MgPc$ is much stronger when it is indirectly excited by RET from the Q_y state than when it is directly excited by charge injection into the $MgPc$ (Extended Data Fig. 6).

The considerable increase in the Q_x fluorescence (2) indicates that the energetically allowed intramolecular state transition $Q_y \rightarrow Q_x$ is not efficient in an isolated H_2Pc molecule and that the presence of a nearby $MgPc$ promotes it. This can be understood by considering that the Q_x and Q_y states are polarized in the x and y directions^{19,28}, respectively, and this orthogonality most probably suppresses the $Q_y \rightarrow Q_x$ transition in isolated H_2Pc . On the NaCl substrate, the polarization of the $MgPc$ Q state is not orthogonal to either the Q_x or Q_y state because H_2Pc and $MgPc$ have different adsorption orientations. The Q state of $MgPc$ is therefore intermediate in terms of both polarization and energy between the Q_x and Q_y states of H_2Pc , and it can act as a mediator to promote the otherwise inefficient $Q_y \rightarrow Q_x$ transition.

Finally, the blinking behaviour of RET from Q to Q_x was analysed³⁰. Figure 4a shows two STL spectra with completely different spectral shapes measured with the same V_t , I_t and tip position. Although spectrum 1 (the blue curve) shows almost no Q_x fluorescence, spectrum 2 (red) shows a Q_x fluorescence peak that is stronger than the Q fluorescence. The components of the Q_x and Q fluorescence were plotted as a function of time (Fig. 4b), showing an anti-correlated blinking of the Q_x and Q fluorescence.

This blinking can be explained by a change in the RET efficiency. If the RET efficiency from Q to Q_x is high, Q fluorescence decreases and Q_x fluorescence increases, and vice versa. The transition rate between the high- and low-RET states is approximately on the order of 1 min^{-1} , which is slow enough for measurement by STL and allowed us to identify the molecular configurations that are associated with the high- and low-RET states (Fig. 4c). In the high-RET state, the x axis of the H_2Pc is directed towards the $MgPc$, and thus the polarization of the Q_x state is almost aligned with the major axis of the $MgPc$ - H_2Pc dimer. In the low-RET state, the Q_x state is polarized almost perpendicular to the major axis. H_2Pc thus essentially acts as a valve device that can control the energy transfer in molecular systems (Fig. 4d).

In closing, we note that the lifetimes of the molecular excited states in the tunnelling junction of only about 10^{-13} s (Extended Data Fig. 7)¹² are much shorter than those observed in conventional experiments in solution. We nevertheless successfully detected RET in our system because its rate was accelerated up to around 10^{13} s^{-1} , achieved by the short ($<2.4 \text{ nm}$) intermolecular distances involved. The challenge now is to explore whether the local excitation dynamics revealed here can be used to realize ultrafast information transfer and processing based on molecular excitonic circuits.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 June; accepted 11 August 2016.

Published online 3 October 2016.

- McDermott, G. *et al.* Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature* **374**, 517–521 (1995).
- Brixner, T. *et al.* Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature* **434**, 625–628 (2005).

- Collini, E. *et al.* Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature* **463**, 644–647 (2010).
- Engel, G. S. *et al.* Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature* **446**, 782–786 (2007).
- Linsebigler, A. L., Lu, G. & Yates, J. T. Photocatalysis on TiO_2 surfaces: principles, mechanisms, and selected results. *Chem. Rev.* **95**, 735–758 (1995).
- O'Regan, B. & Gratzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. *Nature* **353**, 737–740 (1991).
- Tang, C. W. Two-layer organic photovoltaic cell. *Appl. Phys. Lett.* **48**, 183–185 (1986).
- Gimzewski, J. K., Reihl, B., Coombs, J. H. & Schlittler, R. R. Photon emission with the scanning tunneling microscope. *Z. Phys. B* **72**, 497–501 (1988).
- Berndt, R., Gimzewski, J. K. & Johansson, P. Inelastic tunneling excitation of tip-induced plasmon modes on noble-metal surfaces. *Phys. Rev. Lett.* **67**, 3796 (1991).
- Berndt, R. *et al.* Photon-emission at molecular resolution induced by a scanning tunneling microscope. *Science* **262**, 1425–1427 (1993).
- Qiu, X. H., Nazin, G. V. & Ho, W. Vibrationally resolved fluorescence excited with submolecular precision. *Science* **299**, 542–546 (2003).
- Wu, S. W., Nazin, G. V. & Ho, W. Intramolecular photon emission from a single molecule in a scanning tunneling microscope. *Phys. Rev. B* **77**, 205430 (2008).
- Dong, Z. C. *et al.* Generation of molecular hot electroluminescence by resonant nanocavity plasmons. *Nat. Photon.* **4**, 50–54 (2010).
- Chen, C., Chu, P., Bobisch, C. A., Mills, D. L. & Ho, W. Viewing the Interior of a Single Molecule: Vibrationally Resolved Photon Imaging at Submolecular Resolution. *Phys. Rev. Lett.* **105**, 217402 (2010).
- Imada, H. *et al.* Atomic-scale luminescence measurement and theoretical analysis unveiling electron energy dissipation at a p -type GaAs(110) surface. *Nanotechnology* **26**, 365402 (2015).
- Merino, P., Groszke, C., Roslowska, A., Kuhnke, K. & Kern, K. Exciton dynamics of C_{60} -based single-photon emitters explored by Hanbury Brown–Twiss scanning tunnelling microscopy. *Nat. Commun.* **6**, 8461 (2015).
- Zhang, Y. *et al.* Visualizing coherent intermolecular dipole–dipole coupling in real space. *Nature* **531**, 623–627 (2016).
- Chong, M. C. *et al.* Narrow-line single-molecule transducer between electronic circuits and surface plasmons. *Phys. Rev. Lett.* **116**, 036802 (2016).
- Imada, H. *et al.* Orbital-selective single molecule excitation and spectroscopy based on plasmon-exciton coupling. Preprint at <http://arxiv.org/abs/1609.02701> (2016).
- Miwa, K., Sakaue, M. & Kasai, H. Effects of interference between energy absorption processes of molecule and surface plasmons on light emission induced by scanning tunneling microscopy. *J. Phys. Soc. Jpn* **82**, 124707 (2013).
- Miwa, K., Sakaue, M., Gumhalter, B. & Kasai, H. Effects of plasmon energetics on light emission induced by scanning tunneling microscopy. *J. Phys. Condens. Matter* **26**, 222001 (2014).
- Liljeroth, P., Repp, J. & Meyer, G. Current-induced hydrogen tautomerization and conductance switching of naphthalocyanine molecules. *Science* **317**, 1203–1206 (2007).
- Vincett, P. S., Voigt, E. M. & Rieckhoff, K. E. Phosphorescence and fluorescence of phthalocyanines. *J. Chem. Phys.* **55**, 4131–4140 (1971).
- Bartels, L., Meyer, G. & Rieder, K.-H. Controlled vertical manipulation of single CO molecules with the scanning tunneling microscope: a route to chemical contrast. *Appl. Phys. Lett.* **71**, 213–215 (1997).
- Gross, L. Recent advances in submolecular resolution with scanning probe microscopy. *Nat. Chem.* **3**, 273–278 (2011).
- Miwa, K., Imada, H., Kawahara, S. & Kim, Y. Effects of molecule–insulator interaction on geometric property of a single phthalocyanine molecule adsorbed on an ultrathin NaCl film. *Phys. Rev. B* **93**, 165419 (2016).
- Uhlmann, C., Swart, I. & Repp, J. Controlling the orbital sequence in individual Cu-phthalocyanine molecules. *Nano Lett.* **13**, 777–780 (2013).
- Fukuda, R., Ehara, M. & Nakatsuji, H. Excited states and electronic spectra of extended tetraazaporphyrins. *J. Chem. Phys.* **133**, 144316 (2010).
- Repp, J., Meyer, G., Stojković, S. M., Gourdon, A. & Joachim, C. Molecules on insulating films: scanning-tunneling microscopy imaging of individual molecular orbitals. *Phys. Rev. Lett.* **94**, 026803 (2005).
- Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**, 355–358 (1997).

Acknowledgements This work was supported in part by MEXT/JSPS KAKENHI (Grant No. 15H02025, 26886013, 16K21623), and MEXT/JSPS Fellows (No. 15J03915). Some of the numerical computations were performed using RICC and HOKUSAI systems at RIKEN. We thank M. Trenary, H. Kuramochi, K. Inoue and H. Walen for helpful discussions.

Author Contributions H.I., M.I.-I., S.K. and K.K. designed the experiments. H.I. performed the experiment and analysed the data. K.M. provided the theoretical analysis. Y.K. directed the project. All authors discussed the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.K. (ykim@riken.jp).

Reviewer Information Nature thanks L. Gross, G. Nazin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

STM/STS observations. All of the experiments were performed using a low-temperature STM (Omicron) operating at 4.6 K under ultrahigh vacuum. Differential conductance (dI/dV) spectra were measured using a standard lock-in technique with a bias modulation of 20 mV at 617 Hz with an open feedback loop.

Preparation of the sample and tip. The Ag(111) surface was cleaned by repeated cycles of Ar^+ ion sputtering and annealing. The deposition of NaCl onto the Ag(111) at room temperature was performed using a home-made evaporator heated to 850 K. H_2Pc and MgPc were deposited onto the NaCl-covered Ag(111) at 4.7–10 K in the STM head using a commercial three-cell evaporator (Kentax) heated to 575 K for H_2Pc and 625 K for MgPc . CO gas was also introduced on the surface for the tip functionalization. The STM tips were prepared by the electrochemical etching of a Ag wire and conditioned by controlled indentation and voltage pulse on the Ag(111) surface.

STL measurement. The STM stage was designed to be equipped with two optical lenses (each covered a solid angle of ~ 0.5 sr). The emitted light was collimated using the lens and directed out of the ultrahigh vacuum chamber, where it was refocused onto a grating spectrometer (Acton, SpectraPro 2300i) with a charge coupled device photon detector (Princeton, Spec10) cooled with liquid nitrogen. All of the optical spectra (except for Fig. 4, Extended Data Figs 2, 4 and 5) were measured using a grating with 300 grooves per millimetre. For the measurement of spectra shown in Fig. 4 and Extended Data Figs 2, 4 and 5, a lower-energy-resolution grating (50 grooves per millimetre) was used to increase the signal-to-noise ratio and the time response of the STL measurement. The STL spectra are not corrected for the optical throughput of the detection system.

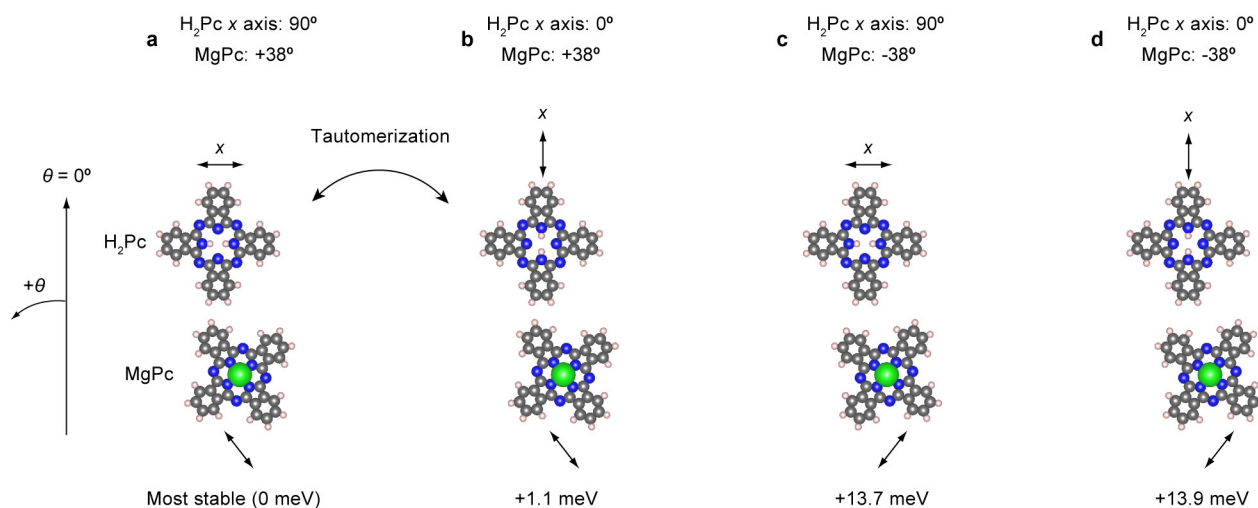
Formation of heterogeneous molecular dimers. First, MgPc and H_2Pc , which are located close to each other on the NaCl film, are found by STM imaging. Then the STM tip is used to change the relative positions of the two molecules slightly. We first found the (5.5, 2.5) dimer for the example in Fig. 1 and the measurements were performed. After the measurement, the H_2Pc was moved by applying a bias

voltage pulse ($V = -3.5$ V) to change the relative position with respect to the MgPc , resulting in the (3.5, 2.5) dimer.

Lifetime estimation. In Extended Data Fig. 7, we estimated the lifetime of the excited states following the method used in a previous STL study of a single molecule¹². It was reported in the previous work that the lifetime of an excited state of magnesium porphyrin (MgP) in the STM junction was on the order of 220 fs, which was determined from the linewidth of STL fluorescence spectrum of MgP as measured by Lorentzian fitting.

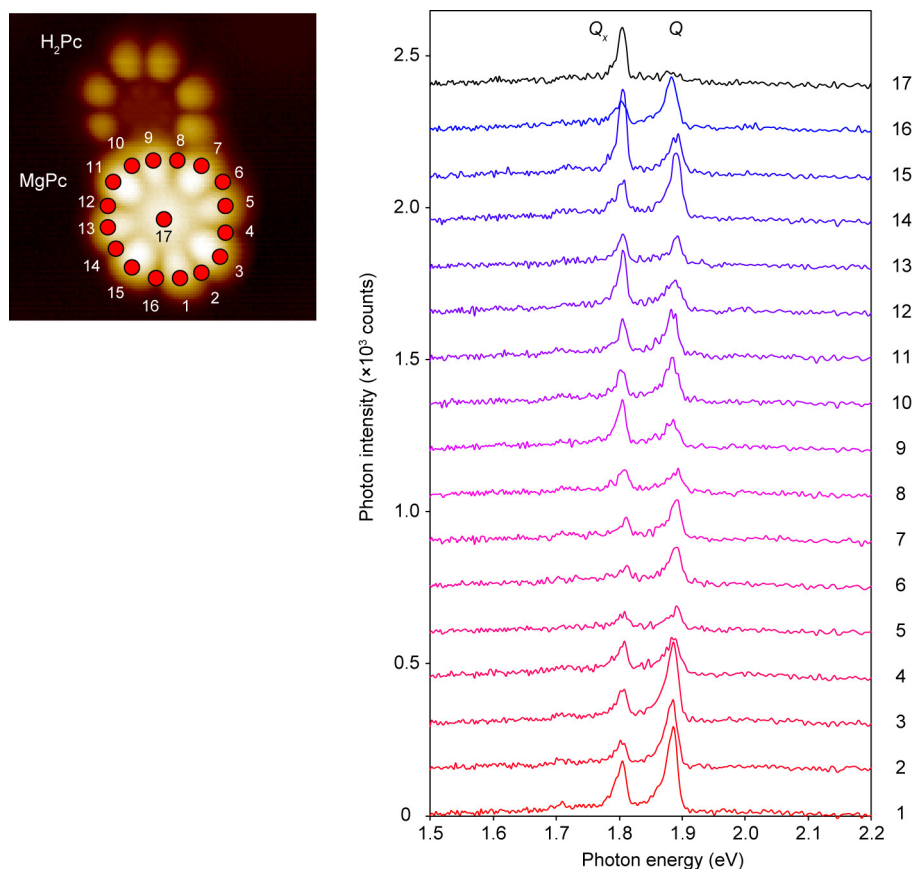
Theoretical calculations. The electronic and geometric structures of H_2Pc and MgPc were investigated using DFT as implemented in the Vienna *Ab initio* Simulation Package code^{31,32}. Generalized gradient approximation within Perdew–Burke–Ernzerhof functional was used to deal with the exchange and correlation effects³³. The interactions between the electron and the ion core were described by the projector augmented wave method³⁴. The one-electron valence states were expanded in a plane-wave basis set with a kinetic-energy cutoff of 480 eV and the Brillouin zone was sampled at the Γ point. Some of the figures were visualized using the Visualization for Electronic and Structural Analysis software³⁵.

31. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
32. Kresse, G. & Furthmüller, J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
33. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
34. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
35. Momma, K. & Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Cryst.* **44**, 1272–1276 (2011).



Extended Data Figure 1 | DFT structural analysis of four possible (3.5, 2.5) MgPc–H₂Pc dimer configurations. **a**, DFT analysis of the total energy indicates that the structure in **a** is the most stable configuration, where the *x* axis of H₂Pc is almost perpendicular to the major axis of the dimer, and MgPc is tilted towards the H₂Pc. **b**, In the second stable configuration, the only difference from **a** is the direction of the *x* axis of the H₂Pc, which is almost parallel to the major axis. The energy difference

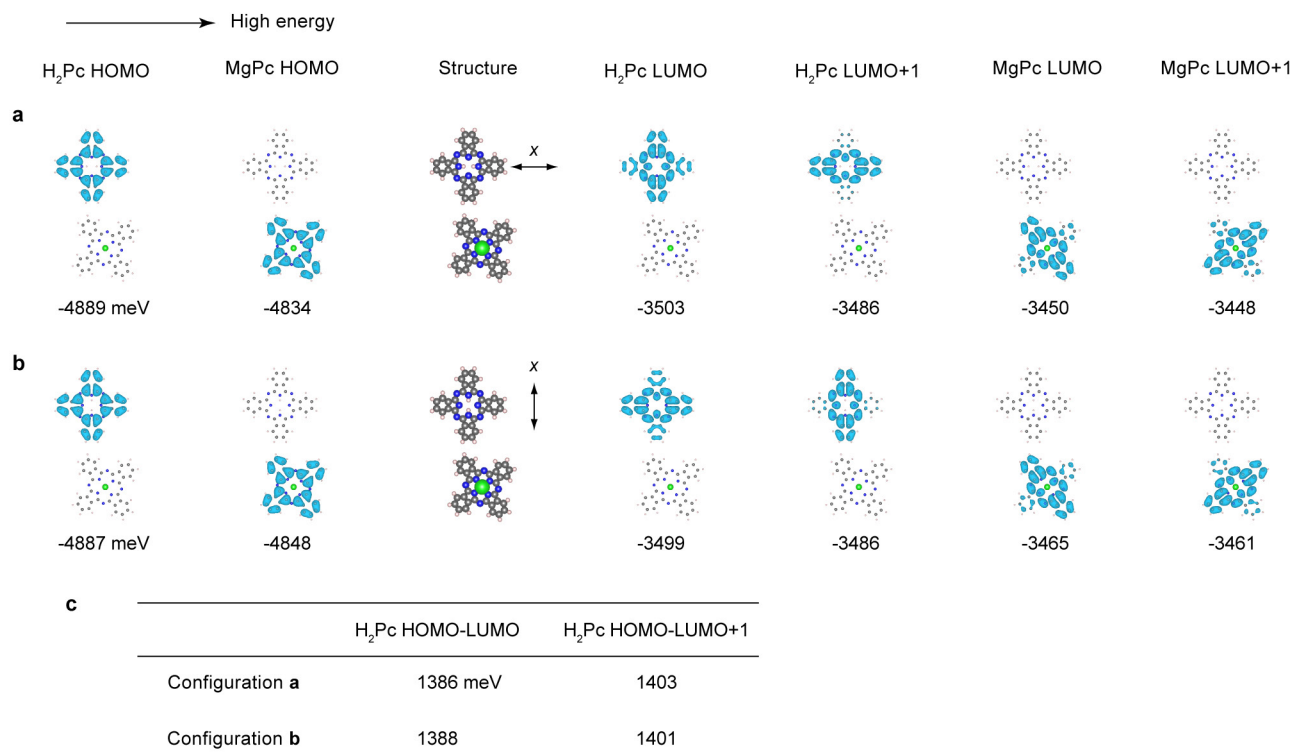
between configurations **a** and **b** is only 1.1 meV, and the tautomerization of the H₂Pc is induced by the tunnelling electron²². **c**, **d**, In configurations **c** and **d** MgPc has different tilt angles, and the total energies increased by around 14 meV, indicating that the two angles $\pm 38^\circ$ are no longer equivalent in the dimer configuration, thus suppressing the shuttling motion of MgPc/NaCl.



Extended Data Figure 2 | Tip position dependence of energy transfer.

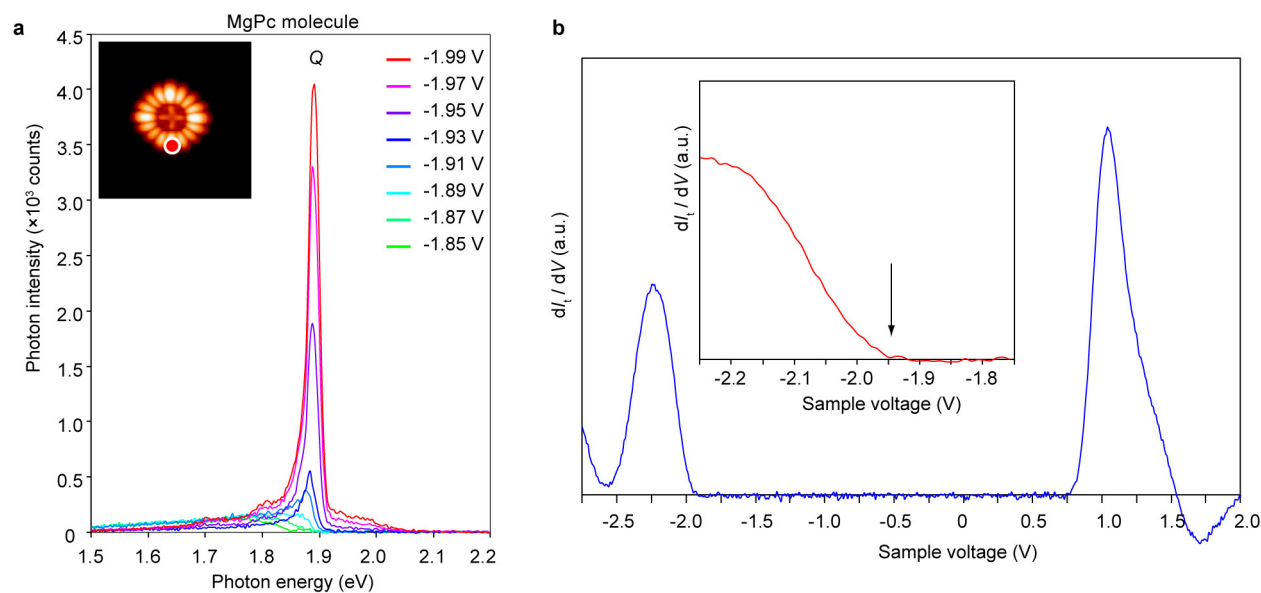
The STL spectra were measured on the MgPc in the (3.5, 2.5) MgPc-H₂Pc dimer at various tip positions ($V = -2.1$ V, $I_t = 5$ pA, $t = 1$ min). The measurement positions are displayed in the STM image. The Q_x line arising from the energy transfer was observed at all points over MgPc, indicating that the energy transfer is not sensitive to the tip position. The Q fluorescence of MgPc was also observed at almost every point, the only exception being at the molecular centre (tip position 17). The Q fluorescence disappeared when the tip was placed at the centre of the MgPc, but the H₂Pc Q_x fluorescence was clearly observed. The suppression of the single-molecule STL when the tip is placed at the molecular centre was reported in previous works^{14,17}. The appearance

of the H₂Pc luminescence when the tip is at the MgPc molecular centre is explained as follows. First, the Q state is excited by charge injection. Although the Q state cannot emit a far-field photon efficiently owing to the STL suppression, the state can transfer its energy to the nearby H₂Pc where plasmon-exciton coupling is allowed, and the H₂Pc exhibits Q_x fluorescence. When the tip is off-centre on the MgPc (positions 1–16), plasmon-exciton coupling is allowed for both MgPc and H₂Pc, which causes the Q and Q_x lines to appear. It should be noted that the blinking behaviour of RET (Fig. 4) makes it difficult to precisely analyse the tip position dependence of the RET probability in our system. The quantitative analysis of the position dependence of RET will be realized with a rigid molecular system.



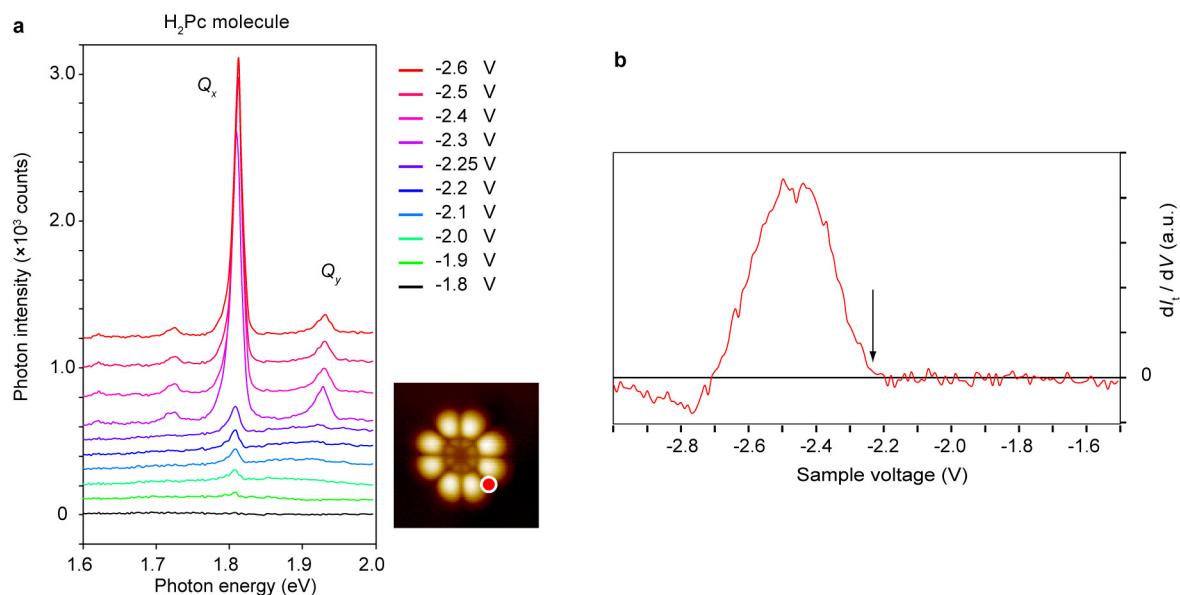
Extended Data Figure 3 | DFT electronic structure analysis of the (3.5, 2.5) MgPc–H₂Pc dimer. **a, b**, Calculated frontier molecular orbitals of the most stable and second most stable structures of the (3.5, 2.5) MgPc–H₂Pc dimer. The isosurface of charge density $|\psi|^2$ and the energy level of each orbital are presented. All of the molecular orbitals are localized in one of the molecules, and no clear hybridization was observed between the orbitals. However, their energy levels were slightly altered by intermolecular interactions. **c**, The energy gaps between the molecular

orbitals at the ground state are listed. Note that the calculated energy gaps cannot be directly compared with the experimental results (Fig. 3a), because the experimentally measured peak positions correspond to energy gaps between the excited state and the ground state. However, the DFT analysis shows that energy levels of the molecular orbitals at the ground state are different in configurations **a** and **b**, suggesting that the resonance energies of the electronic transitions among them are also different.



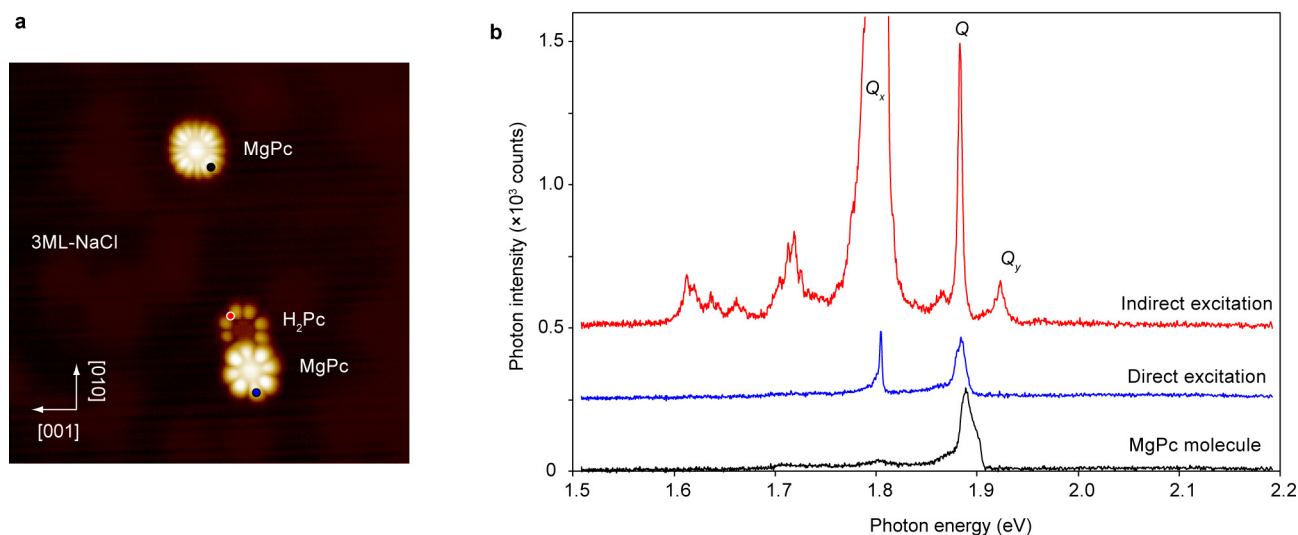
Extended Data Figure 4 | Determination of the threshold voltage required to induce single-molecule fluorescence of the MgPc molecule. **a**, The bias voltage-dependent STL spectra of MgPc/3ML NaCl were measured with $I_t = 20$ pA, $t = 1$ min, at the red dot in the inset. This shows that the threshold voltage was -1.95 V. **b**, A dI_t/dV spectrum of

MgPc measured with the same tip used in **a** (at the red dot in the inset of **a**). As reported previously, the threshold voltage for single-molecule electroluminescence corresponds to that of the resonant tunnelling channel through the HOMO in dI_t/dV spectrum¹⁷.



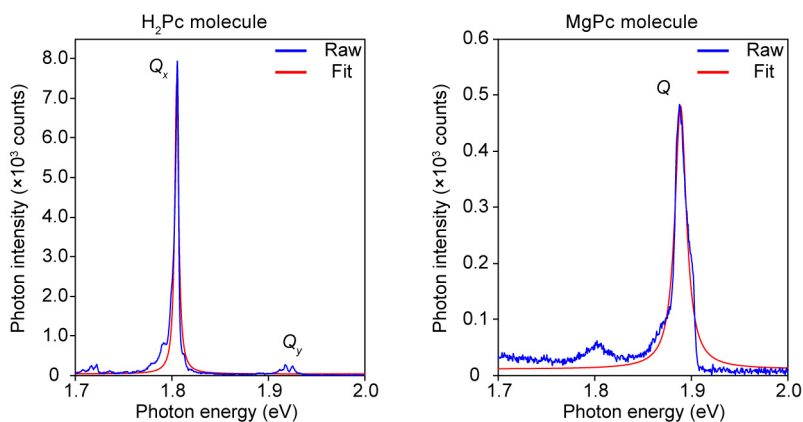
Extended Data Figure 5 | Threshold voltages to induce single-molecular fluorescence of the H₂Pc molecule. **a**, The bias voltage-dependent STPL spectra of H₂Pc/3ML NaCl were measured with $I_t = 25$ pA and $t = 1$ s, at the red dot in the inset. **b**, A dI_t/dV spectrum of H₂Pc in the threshold voltage region. When $|V| \leq 1.8$ V, the spectrum shows only the radiation of the localized plasmon (the intensity is very weak) and no molecular fluorescence was detected. When 1.8 V $< |V| \leq 2.25$ V, weak Q_x fluorescence appeared at 1.81 eV and Q_y fluorescence was barely seen. When 2.25 V $< |V|$, strong Q_x fluorescence and weak Q_y fluorescence were observed. It is clear that there are two threshold voltages, $V_{th1} = -1.8$ V for

Q_x fluorescence and $V_{th2} = -2.3$ V for both Q_x and Q_y fluorescence of H₂Pc. V_{th1} corresponds to the energy of the Q_x state (1.81 eV), and V_{th2} to the threshold voltage of the resonant tunnelling channel through the HOMO in dI_t/dV spectrum as seen in **b**. The former is similar to the process reported in ref. 18, and the latter was described in ref. 17. Although another threshold voltage was expected at -1.92 V, which corresponds to the energy of the Q_y state (1.92 eV), it was not clearly observed in our experiment because of the weakness of the Q_y luminescence. The strong single-molecule electroluminescence of H₂Pc is triggered by hole injection into the HOMO, which is similar to the case of MgPc.



Extended Data Figure 6 | Direct and indirect excitation of the Q fluorescence of MgPc. **a**, An STM image of the (3.5, 2.5) MgPc-H₂Pc dimer and an MgPc molecule ($V = -2.3$ V, $I_t = 5$ pA). **b**, The STL spectra measured at the three different positions indicated in **a** were compared. The red, blue and black curves were measured at the red, blue and black points in **a**, respectively, with the same measurement parameters ($V = -2.3$ V, $I_t = 30$ pA, $t = 1$ min). The integrated photon intensities in the range 1.871–1.908 eV were 13,000, 5,116 and 8,823 counts for the red, blue,

and black curves, respectively. The results clearly show that the excitation of the Q state of MgPc is much more efficient when induced by indirect excitation through RET from the Q_y state of the nearby H₂Pc than by direct excitation with the tunnelling current. It is therefore concluded that the main excitation mechanism of the Q state is RET from the Q_y state under the measurement conditions of the red curve (which is the same spectrum as the red curve shown in Fig. 3).



Extended Data Figure 7 | Lifetime estimation from the linewidths.

The linewidths observed in our experiment were 4.8 meV for the H₂Pc Q_x fluorescence and 14.4 meV for the MgPc Q fluorescence measured by single Lorentzian fitting. The linewidth of the MgPc Q fluorescence is not determined only by the lifetime of the state, because the line shape is not a simple Lorentz function and it is possible that other radiative processes are involved in the peak. In contrast, the H₂Pc Q_x fluorescence is reasonably fitted with a single Lorentz function, suggesting that the linewidth is mostly determined by the lifetime of the Q_x state. The 4.8 meV

linewidth is similar to the previously reported value (4.4 meV; ref. 12), and the lifetime of the Q_x state is estimated to be approximately a few hundred femtoseconds (about 10^{−13} s). We believe that the Q state of MgPc also has a similar lifetime, because the magnitudes of the transition dipole moments and the spatial distributions of the molecular orbitals are similar for H₂Pc and MgPc. The difference in the STL line shape might arise from different vibrational interactions with the NaCl substrate, which may be expected from the very different adsorption configurations of the two molecules.

Asthenosphere rheology inferred from observations of the 2012 Indian Ocean earthquake

Yan Hu^{1,2}, Roland Bürgmann¹, Paramesh Banerjee³, Lujia Feng³, Emma M. Hill³, Takeo Ito⁴, Takao Tabei⁵ & Kelin Wang⁶

The concept of a weak asthenospheric layer underlying Earth's mobile tectonic plates is fundamental to our understanding of mantle convection and plate tectonics. However, little is known about the mechanical properties of the asthenosphere (the part of the upper mantle below the lithosphere) underlying the oceanic crust, which covers about 60 per cent of Earth's surface. Great earthquakes cause large coseismic crustal deformation in areas hundreds of kilometres away from and below the rupture area. Subsequent relaxation of the earthquake-induced stresses in the viscoelastic upper mantle leads to prolonged postseismic crustal deformation that may last several decades and can be recorded with geodetic methods^{1–3}. The observed postseismic deformation helps us to understand the rheological properties of the upper mantle, but so far such measurements have been limited to continental-plate boundary zones. Here we consider the postseismic deformation of the very large (moment magnitude 8.6) 2012 Indian Ocean earthquake^{4–6} to provide by far the most direct constraint on the structure of oceanic mantle rheology. In the first three years after the Indian Ocean earthquake, 37 continuous Global Navigation Satellite Systems stations in the region underwent horizontal northeastward displacements of up to 17 centimetres in a direction similar to that of the coseismic offsets. However, a few stations close to the rupture area that had experienced subsidence of up to about 4 centimetres during the earthquake rose by nearly 7 centimetres after the earthquake. Our three-dimensional viscoelastic finite-element models of the post-earthquake deformation show that a thin (30–200 kilometres), low-viscosity (having a steady-state Maxwell viscosity of $(0.5\text{--}10) \times 10^{18}$ pascal seconds) asthenospheric layer beneath the elastic oceanic lithosphere is required to produce the observed postseismic uplift.

We analysed the time series recorded by 47 continuous Global Navigation Satellite Systems (GNSS) stations, including 31 from the Sumatran Global Positioning System (GPS) Array (SuGAR), 11 from the International GNSS Service (IGS), 3 from the University of Memphis Andaman Island network, and 2 from the Aceh GPS Network for the Sumatran Fault System (AGNeSS). We selected 37 of these stations, those that show a coherent pattern of postseismic motions and do not have data gaps during the Indian Ocean earthquake (IOE) (Extended Data Fig. 1). The IOE produced static coseismic offsets of more than 20 cm at stations less than 500 km from the rupture area^{7,8} and subsidence of up to about 4 cm (Fig. 1a). After removing the effects of previous earthquakes and the coseismic offsets of the IOE, as well as secular, annual and semi-annual trends (Extended Data Figs 2, 3)⁹, we derived postseismic displacements of these stations in the first 3 years following the IOE. We find horizontal motion of up to about 17 cm in a landward direction similar to that of the coseismic displacements (Fig. 1b). The striking feature of the postseismic vertical displacement is that these middle-field stations within 300–500 km of the mainshock have risen by up to about 7 cm, reversing the coseismic subsidence,

which is consistent with reported positive postseismic gravity changes in the same area¹⁰.

On the basis of previous studies of subduction zone earthquakes in Sumatra^{11,12} and other convergent margins^{2,13,14}, we constructed a viscoelastic finite-element model invoking the biviscous Burgers rheology¹⁵ (Fig. 2) to study the postseismic deformation of the IOE. Transient Kelvin viscosity η_K is assumed to be one order of magnitude lower than the steady-state Maxwell viscosity η_M (the viscosity hereafter in this paper refers to the steady-state viscosity unless explicitly stated otherwise). Given the limited timespan of the GNSS data, we thus provide a lower-bound estimate of the steady-state viscosities.

The IOE involved a composite rupture of six strike-slip faults. Postseismic deformation at GNSS stations hundreds of kilometres from the rupture area is sensitive to the total moment of the earthquake, not to details of the slip distribution. Different coseismic fault slip models^{6,7,16} predict different patterns of near-field postseismic displacements within 300 km of the mainshock but almost identical displacements at the GNSS stations (Extended Data Fig. 4). The coseismic fault slip distribution determined by Wei *et al.*⁶ is used in this work.

We examine a number of first-order model scenarios to motivate our choice of primary model parameters, which we then evaluate in more detail. Assuming only one homogeneous viscoelastic layer below the elastic lithosphere, we need to use a low viscosity in the oceanic upper mantle of order 10^{19} Pa s to fit the observed horizontal GNSS data (Extended Data Fig. 5b). However, this test model results in postseismic subsidence that is inconsistent with the observed GNSS uplift. We find that models including a thin low-viscosity top layer of the oceanic asthenosphere can readily produce the observed uplift. Varying the lithospheric thickness by 20 km (Extended Data Fig. 6a, b) or imposing a smooth gradient in viscosity at the lithosphere–asthenosphere boundary (Extended Data Fig. 9c) produces negligible changes in the postseismic motions at GNSS stations. However, the effects of the subducting slab cannot be ignored (Extended Data Fig. 6c).

We assume the viscosity of the mantle wedge overlying the subducting Indo-Australian plate to be 3×10^{19} Pa s (ref. 13), but changing this value by one order of magnitude has little effect on predicted postseismic displacements at our GNSS stations. The postseismic surface deformation is controlled mainly by the rheological structure of the oceanic upper mantle (Extended Data Fig. 7). The rheological properties of the oceanic asthenosphere and upper mantle obtained in this work are better resolved at depths of less than 400 km because the IOE-induced stresses at greater depths are negligibly small (results not shown).

We use a grid-search method to determine preferred values of three model parameters from hundreds of models: the thickness (D_A) and viscosity (η_A) of the oceanic asthenosphere and the viscosity of the underlying oceanic upper mantle (η_O). We vary D_A , η_A and η_O within the ranges 10–300 km, 10^{17} – 10^{20} Pa s and 10^{19} – 10^{22} Pa s, respectively. To find the best-fit model parameters and their tradeoffs, we calculate

¹Berkeley Seismological Laboratory and Department of Earth and Planetary Science, University of California Berkeley, Berkeley, California, USA. ²Mengcheng National Geophysical Observatory, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China. ³Earth Observatory of Singapore, Asian School of the Environment, Nanyang Technological University, Singapore. ⁴Graduate School of Environmental Studies, Nagoya University, Aichi 464-8601, Japan. ⁵Department of Applied Science, Kochi University, Akebono-cho 2-5-1, Kochi 780-8520, Japan. ⁶Pacific Geoscience Centre, Geological Survey of Canada, Natural Resources Canada, Sidney, British Columbia, Canada.

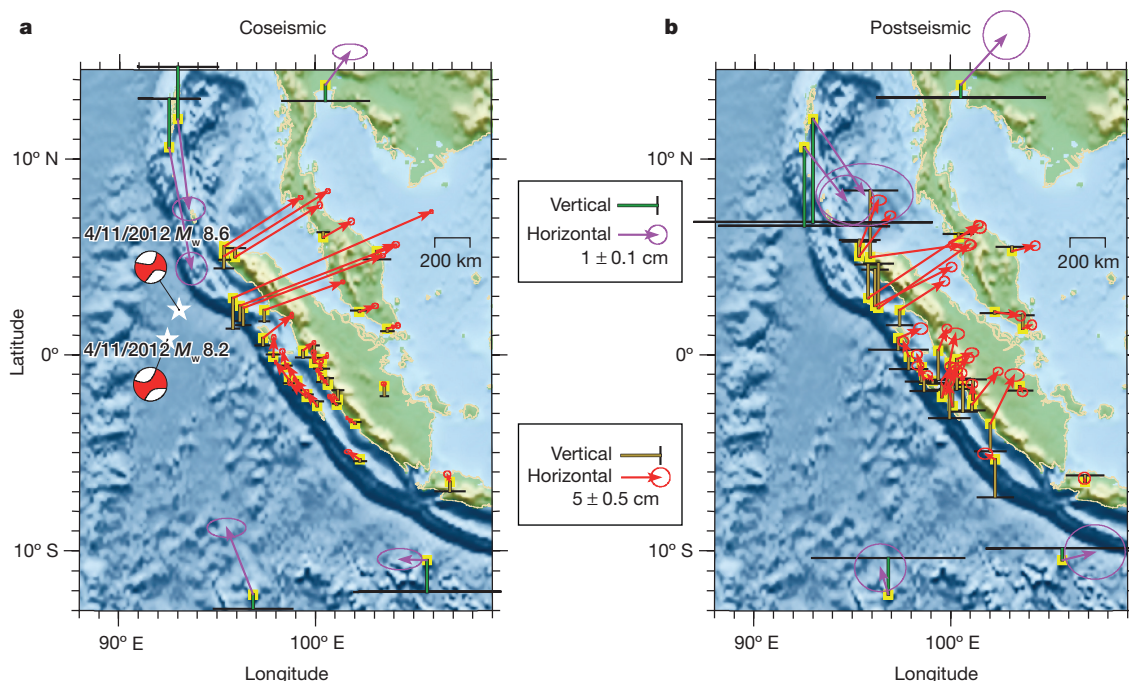


Figure 1 | Coseismic and cumulative three-year-postseismic GNSS observations of the IOE. Error bars represent 2σ (95%) confidence intervals. Yellow squares represent locations of the GNSS stations. **a**, Coseismic displacements of the IOE, estimated from static offsets of five days before and after the IOE (the IOE includes two events—the

two beachballs—separated by about two hours). **b**, Cumulative three-year-postseismic displacements of the IOE in the Sunda reference frame (Supplementary Table 1). Red and magenta arrows represent horizontal GNSS displacements at different scales. Brown and green bars represent vertical GNSS displacements at different scales.

the χ^2 misfit of each test model prediction (equation (1) in Methods) to our GNSS displacements.

If we consider χ^2 only in the horizontal components (Fig. 3a), a test model fitting to the GNSS observations requires $D_A \geq 50$ km, η_A of the order of 10^{19} Pa s and $\eta_O \geq 10^{19}$ Pa s. The test model that best fits the horizontal GNSS motion does not predict the observed forearc uplift (Extended Data Fig. 9d). If we consider χ^2 only in the vertical component (Fig. 3b), a η_A value of the order of 10^{18} Pa s produces a good fit to the vertical GNSS displacements. The test model that best fits the vertical GNSS motion overestimates the horizontal components in the middle field (Extended Data Fig. 9e).

If we consider χ^2 in both the horizontal and vertical components (Fig. 3c), all three model parameters are constrained within a relatively narrow range. D_A , η_A and η_O are determined to be in the ranges 30–200 km, $(0.5\text{--}10) \times 10^{18}$ Pa s and $(0.5\text{--}100) \times 10^{20}$ Pa s, respectively.

The lowest- χ^2 preferred model (PM) has $D_A = 80$ km, $\eta_A = 2 \times 10^{18}$ Pa s, and $\eta_O = 10^{20}$ Pa s (Extended Data Fig. 9f). The first-order mantle structure obtained in this work is consistent with results from a regional surface-wave tomography study¹⁷ that indicates a low-velocity region centred at a depth of about 150 km.

There are important tradeoffs between model parameters, especially between the thickness and viscosity of the asthenospheric layer. If η_O is fixed at 10^{20} Pa s as in the PM, η_A scales with D_A because $\eta_A = aD_A^{1.5}$, where $a = 3.5 \times 10^{15}$ Pa s km^{-1.5} (Fig. 4a), and D_A is in kilometres. This tradeoff is similar to the one found in models of isostatic rebound of continental regions that were covered by thick ice caps during the last ice age. Paulson *et al.*¹⁸ analysed the postglacial rebound relying on long-wavelength (>700 km) Gravity Recovery And Climate Experiment (GRACE) satellite data in Canada and the sea-level history in Hudson Bay and reported a similar relationship, $\eta_A \propto D_A^3$.

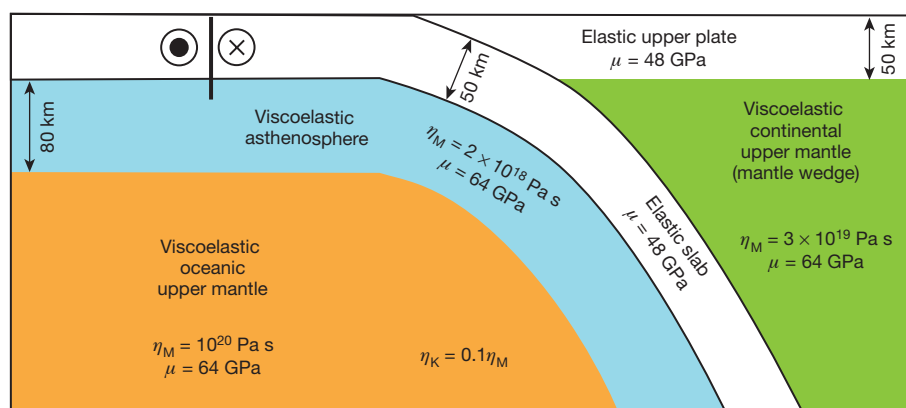


Figure 2 | Conceptual representation of the finite-element model. The model includes an elastic upper plate and elastic slab, viscoelastic continental upper mantle (mantle wedge), viscoelastic oceanic asthenosphere and viscoelastic oceanic upper mantle. The rock

properties of each structural unit are given: μ , η_M and η_K represent the shear modulus, steady-state Maxwell and transient Kelvin viscosities, respectively. $\eta_K = 0.1\eta_M$. The thick black line illustrates the strike-slip fault of the IOE.

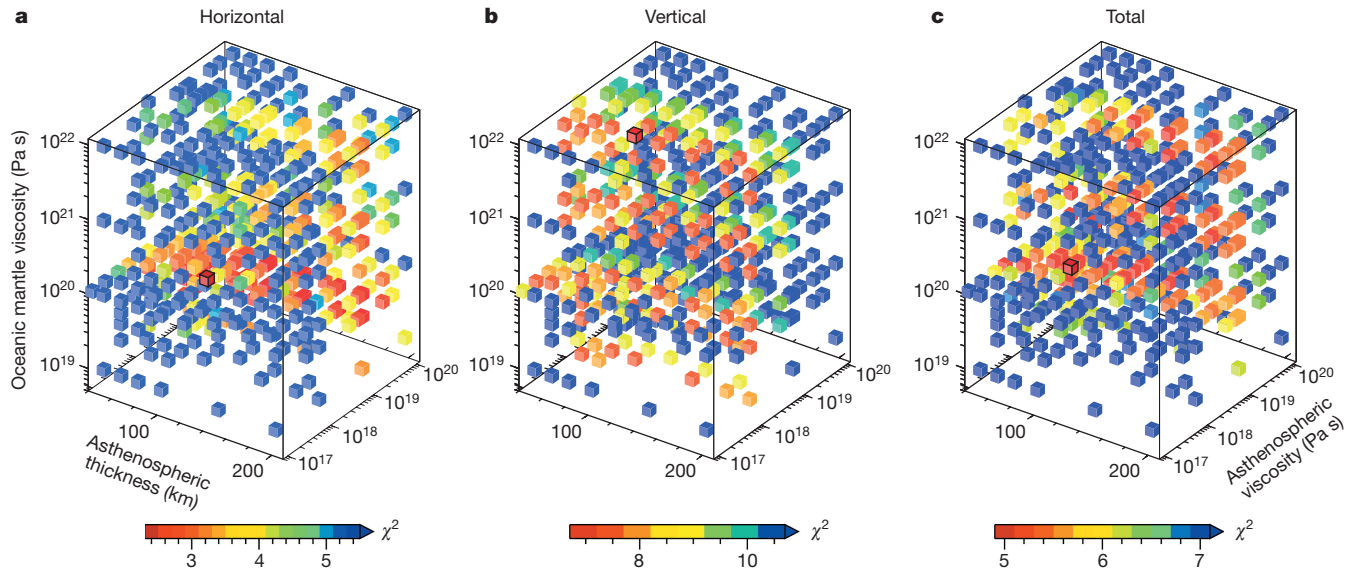


Figure 3 | Misfit of 652 test models considering variations in the asthenospheric thickness and viscosity, and oceanic mantle viscosity. Each cube represents one test model. Test models with red colour (low χ^2 values) reproduce the overall pattern of GNSS observations.

Their higher power of D_A may be due to the low spatial resolution of the GRACE data and the much greater lithospheric thickness and higher mantle viscosities of the North American interior. η_O is correlated with D_A and shows a modest anti-correlation with η_A (Fig. 4b, c).

The PM well reproduces the overall magnitude of the observed uplift in the mid-field forearc area (Fig. 5a). The PM also reproduces the first-order pattern of the GNSS observations in the far field more than 500 km from the mainshock. The large misfit at stations between latitudes 0° and 6° S may be due to the low signal-to-noise ratio at those stations. The remaining misfits to the data, including the slight overestimates of the horizontal displacements in the mid-field, may indicate additional complexity of the rheology structure and other local processes, such as aftershocks and aseismic afterslip of the IOE, which are not considered in the PM. The PM predicted displacement evolution also matches the general curvature of the time series of the GNSS stations with three examples shown in Fig. 5b–d. The model predicts that the vertical displacement may soon reverse direction in the continental

area, but not the horizontal components (Extended Data Fig. 10). The vertical component is more sensitive than the horizontal components to the change in the pattern of the viscoelastic flow above and beneath the slab caused by the existence of the elastic slab.

We did not include contributions from aseismic afterslip in the PM. We study the effects of stress-driven afterslip around the rupture segments of the IOE using the approach presented in Hu *et al.*¹³, which relies on 2-km-thick low-viscosity tabular shear zones adjacent to the rupture. The afterslip model, regardless of assumed shear zone viscosity, overestimates the horizontal GNSS displacements (Extended Data Fig. 8a–d). Increasing the viscosity of the asthenosphere can lessen the effect of afterslip on the horizontal motions, but worsens the fit to the vertical GNSS component (Extended Data Fig. 8f). However, we cannot rule out a scenario of deep afterslip at depths of more than 50 km that produces displacements of up to 30 cm three years after the IOE in the near field but negligible motions at GNSS stations (Extended Data Fig. 8e). Nevertheless, substantial afterslip following the IOE, at

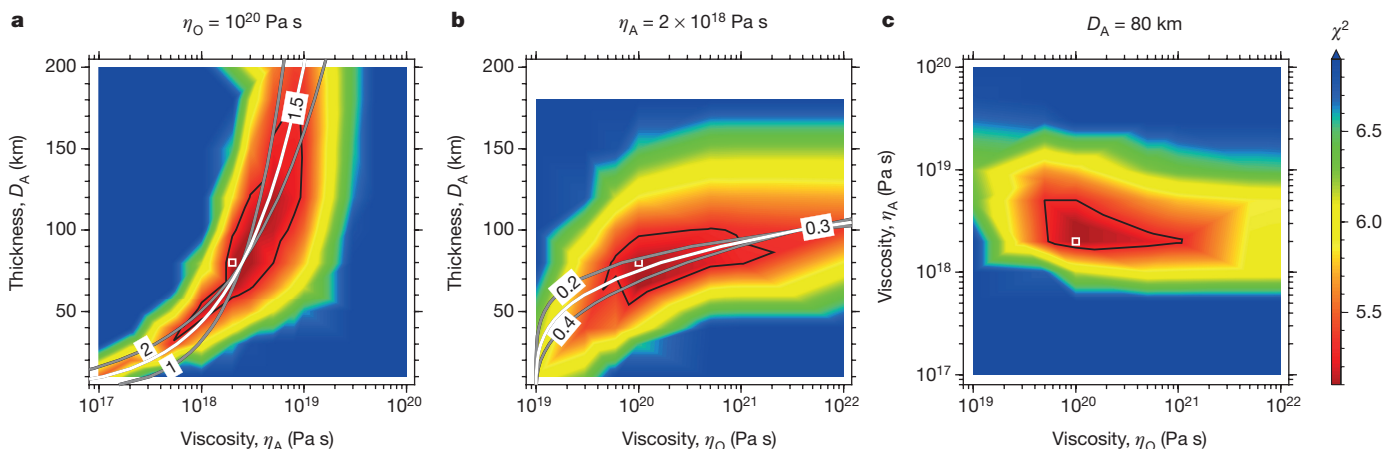


Figure 4 | Tradeoff between the viscosity of the oceanic upper mantle (η_O), the thickness (D_A) of the asthenosphere and its viscosity (η_A). The χ^2 misfit of the models is shown by the colour contours. Solid black lines represent the upper bound of $\chi^2 = 5.3$, below which the models match the overall pattern of the GNSS observations. White squares represent the PM. **a**, η_O is fixed at 10^{20} Pa s. The thick white line represents the preferred

thickness–viscosity tradeoff relationship (with power 1.5): that is, $\eta_A = 3.5 \times 10^{15} D_A^{1.5}$. Grey lines represent different powers (1 and 2) of D_A . **b**, η_A is fixed at 2×10^{18} Pa s. The thick white line represents the preferred D_A – η_O relationship (with power 0.3): that is, $D_A = 75(\log_{10} \eta_O - 19)^{0.3}$. Grey lines represent different powers (0.2 and 0.4). **c**, D_A is fixed at 80 km.

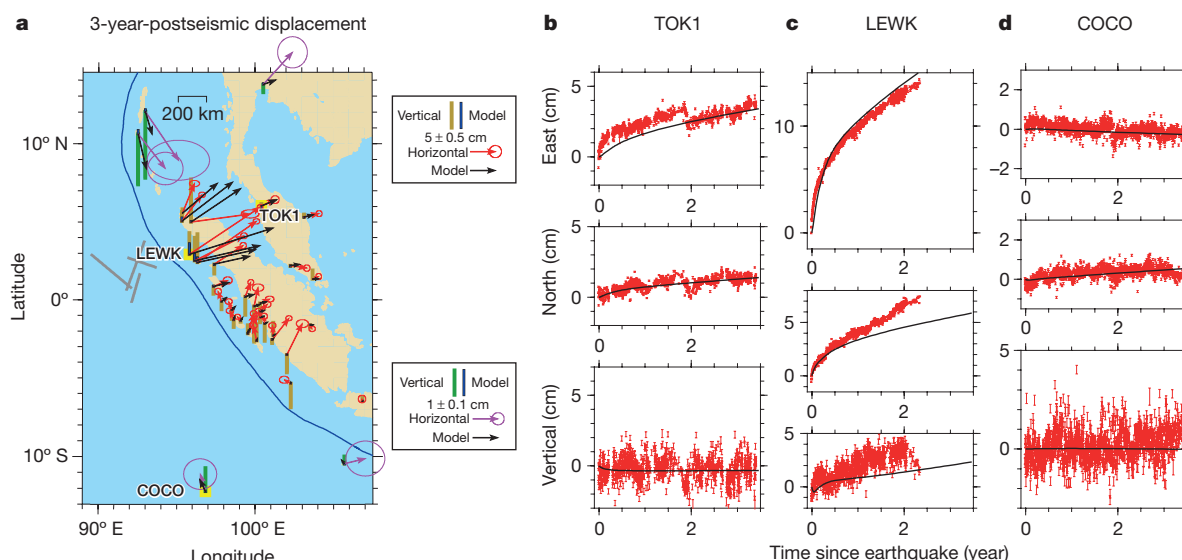


Figure 5 | Comparison of GNSS observations with predictions of the PM. **a**, Cumulative three-year-postseismic displacements. Red and magenta arrows represent observed horizontal GNSS displacements at different scales. Solid green and brown bars represent vertical GNSS observations at different scales. Black arrows and solid blue bars (note that for far-field stations, these are too small to be visible) represent horizontal and vertical model predictions, respectively. Yellow squares represent the

locations of three example GNSS stations whose time series are shown in **b–d**. Thick grey lines represent the rupture segments of the IOE⁶. **b–d**, Comparison of GNSS time series with model-predicted displacements at stations TOK1, LEWK and COCO, respectively. Red dots with error bars indicating the 1σ uncertainties represent the GNSS observations. Black lines show model-predicted displacements.

shallow depths in particular, is unlikely to have occurred, as it would have produced subsidence in the northern Sumatra forearc.

If the asthenospheric layer terminates at the trench, this layer must have a lower viscosity or larger thickness to produce a comparable goodness of fit to the land GNSS data (Extended Data Fig. 9a, b). In the PM the oceanic asthenosphere extends with the subducting slab, based on some seismic imaging studies^{17,19,20} and geodynamic modelling²¹. A denser geodetic network, particularly with near-field seafloor geodetic measurements, and a longer timespan of postseismic observations would help resolve this model ambiguity.

The purpose of this work is to study the first-order approximation of the viscoelastic relaxation of the upper mantle on the postseismic deformation of the 2012 earthquake. Therefore we do not consider a more complex thermal- and pressure-dependent rheology that may better represent the real Earth. Poroelastic rebound in the top layer of the lithosphere caused by the earthquake contributes to the postseismic deformation mainly in the vicinity of the rupture region⁹, and is not considered in this work, which studies only the mid- and far-field deformation.

Improved knowledge of the depth and nature of the oceanic lithosphere–asthenosphere boundary and the rheology of the asthenosphere is essential to understanding the interplay of mantle convection and plate tectonics^{22–24}. A weak asthenosphere lubricates plate tectonics, allows for rapid changes in plate motion, and enables lateral flow of upper-mantle material that produces vertical motions of the seafloor and continental margins^{22,25}. A low-viscosity layer may also promote postseismic strain and stress transients that may affect seismicity rates over long distances and time spans²⁶. A range of seismological and electrical resistivity observations show a sharp change in mantle properties at the boundary, indicating the presence of partial melt or water in the asthenosphere^{27,28}. For example, Naif *et al.*²⁹ analysed sea-floor magnetotelluric data to reveal a partially melted channel less than 30 km thick along the lithosphere–asthenosphere boundary beneath the oceanic lithosphere of the Cocos plate. Stern *et al.*³⁰ relied on seismic reflection data to document a similar layer of approximately 10 km thickness at the base of the Pacific plate, subducting beneath the North Island of New Zealand. Other seismologic and petrological observations also favour a sharp boundary over a relatively thin, partially melted

low-velocity zone^{31–33} that decouples the oceanic lithosphere from the underlying mantle. Although there is a tradeoff between the viscosity and thickness of the low-viscosity layer on the lithosphere of the Indian Ocean, our results confirm the interpretation of the geophysical observations as reflecting the existence of a low-viscosity asthenosphere underlying the oceanic lithosphere.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April; accepted 16 August 2016.

Published online 10 October 2016.

- Wang, K., Hu, Y. & He, J. Deformation cycles of subduction earthquakes in a viscoelastic Earth. *Nature* **484**, 327–332 (2012).
- Hu, Y., Wang, K., He, J., Klotz, J. & Khazaradze, G. Three-dimensional viscoelastic finite element model for post-seismic deformation of the great 1960 Chile earthquake. *J. Geophys. Res.* **109**, B12403 (2004).
- Suito, H. & Freymueller, J. T. A viscoelastic and afterslip postseismic deformation model for the 1964 Alaska earthquake. *J. Geophys. Res.* **114**, B11404 (2009).
- Meng, L. *et al.* Earthquake in a maze: compressional rupture branching during the 2012 M_w 8.6 Sumatra earthquake. *Science* **337**, 724–726 (2012).
- Pollitz, F. F., Stein, R. S., Sevilgen, V. & Bürgmann, R. The 11 April 2012 east Indian Ocean earthquake triggered large aftershocks worldwide. *Nature* **490**, 250–253 (2012).
- Wei, S., Helmberger, D. & Avouac, J.-P. Modeling the 2012 Wharton basin earthquakes off-Sumatra: complete lithospheric failure. *J. Geophys. Res. Solid Earth* **118**, 3592–3609 (2013).
- Hill, E. M. *et al.* The 2012 M_w 8.6 Wharton Basin sequence: a cascade of great earthquakes generated by near-orthogonal, young, oceanic mantle faults. *J. Geophys. Res. Solid Earth* **120**, 3723–3747 (2015).
- Feng, L. *et al.* A unified GPS-based earthquake catalog for the Sumatran plate boundary between 2002 and 2013. *J. Geophys. Res. Solid Earth* **120**, 3566–3598 (2015).
- Hu, Y. *et al.* Contributions of poroelastic rebound and a weak volcanic arc to the postseismic deformation of the 2011 Tohoku earthquake. *Earth Planet Space* **66**, 106 (2014).
- Han, S.-C., Sauber, J. & Pollitz, F. Coseismic compression/dilatation and viscoelastic uplift/subsidence following the 2012 Indian Ocean earthquakes quantified from satellite gravity observations. *Geophys. Res. Lett.* **42**, 3764–3772 (2015).
- Pollitz, F. F., Banerjee, P., Grijalva, K., Nagarajan, B. & Bürgmann, R. B. Effect of 3-D viscoelastic structure on post-seismic relaxation from the 2004 $M = 9.2$ Sumatra earthquake. *J. Int.* **173**, 189–204 (2008).

12. Wiseman, K., Bürgmann, R., Freed, A. M. & Banerjee, P. Viscoelastic relaxation in a heterogeneous Earth following the 2004 Sumatra-Andaman earthquake. *Earth Planet. Sci. Lett.* **431**, 308–317 (2015).
13. Hu, Y. *et al.* Stress-driven relaxation of heterogeneous upper mantle and time-dependent afterslip following the 2011 Tohoku earthquake. *J. Geophys. Res. Solid Earth* **121**, 385–411 (2016).
14. Sun, T. *et al.* Prevalence of viscoelastic relaxation after the 2011 Tohoku-Oki earthquake. *Nature* **514**, 84–87 (2014).
15. Bürgmann, R. & Dresen, G. Rheology of the lower crust and upper mantle: evidence from rock mechanics, geodesy, and field observations. *Annu. Rev. Earth Planet. Sci.* **36**, 531–567 (2008).
16. Yadav, R. K., *et al.* Coseismic offsets due to the 11 April 2012 Indian Ocean earthquakes (Mw 8.6 and 8.2) derived from GPS measurements. *Geophys. Res. Lett.* **40**, 3389–3393 (2013).
17. Shapiro, N. M., Ritzwoller, M. H. & Engdahl, E. R. Structural context of the great Sumatra-Andaman Islands earthquake. *Geophys. Res. Lett.* **35**, L05301 (2008).
18. Paulson, A. & Richards, M. A. On the resolution of radial viscosity structure in modelling long-wavelength postglacial rebound data. *Geophys. J. Int.* **179**(3), 1516–1526 (2009).
19. Song, T.-R. A. & Kawakatsu, H. Subduction of oceanic asthenosphere: evidence from sub-slab seismic anisotropy. *Geophys. Res. Lett.* **39**, L17301 (2012).
20. Huang, J. & Zhao, D. High-resolution mantle tomography of China and surrounding regions. *J. Geophys. Res.* **111**, B09305 (2006).
21. Liu, L. & Zhou, Q. Deep recycling of oceanic asthenosphere material during subduction. *Geophys. Res. Lett.* **42**, 2204–2211 (2015).
22. Bercovici, D., Tackley, P. & Ricard, Y. in *Treatise on Geophysics Vol. 7 Mantle Dynamics* (eds Bercovici, D. & Schubert, G.) 271–318 (Elsevier, 2015).
23. Weismüller, J. *et al.* Fast asthenosphere motion in high-resolution global mantle flow models. *Geophys. Res. Lett.* **42**, 7429–7435 (2015).
24. Höink, T., A. Jellinek, M. & Lenardic, A. Viscous coupling at the lithosphere–asthenosphere boundary. *Geochem. Geophys. Geosyst.* **12**, Q0AK02 (2011).
25. Colli, L. I. *et al.* Rapid South Atlantic spreading changes and coeval vertical motion in surrounding continents: evidence for temporal changes of pressure-driven upper mantle flow. *Tectonics* **32**, 1304–1321 (2014).
26. Pollitz, F. F., Bürgmann, R. & Romanowicz, B. Viscosity of oceanic asthenosphere inferred from remote triggering of earthquakes. *Science* **280**, 1245–1249 (1998).
27. Fischer, K. M., Ford, H. A., Abt, D. L. & Rychert, C. A. The lithosphere–asthenosphere boundary. *Annu. Rev. Earth Planet. Sci.* **38**, 551–575 (2010).
28. Karato, S. On the origin of the asthenosphere. *Earth Planet. Sci. Lett.* **321–322**, 95–103 (2012).
29. Naif, S., Key, K., Constable, S. & Evans, R. L. Melt-rich channel observed at the lithosphere–asthenosphere boundary. *Nature* **495**, 356–359 (2013).
30. Stern, T. A. *et al.* A seismic reflection image for the base of a tectonic plate. *Nature* **518**, 85–88 (2015).
31. Schmerr, N. The Gutenberg discontinuity: melt at the lithosphere–asthenosphere boundary. *Science* **335**, 1480–1483 (2012).
32. Yamamoto, J., Korenaga, J., Hirano, N. & Kagi, H. Melt-rich lithosphere–asthenosphere boundary inferred from petit-spot volcanoes. *Geology* **42**, 967–970 (2014).
33. Kawakatsu, H. *et al.* Seismic evidence for sharp lithosphere–asthenosphere boundaries of oceanic plates. *Science* **324**, 499–502 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by HPC resources from the Arctic Region Supercomputing Center and the University of Alaska Fairbanks. J. He of the Geological Survey of Canada wrote the finite-element computer code used in this work. This work was funded by NSF award EAR-1246850 and benefited from support by the Miller Institute for Basic Research in Science to R.B. and a Singapore National Research Foundation Fellowship to E.M.H. (NRF-NRFF2010-064). J. Paul from the University of Memphis provided GPS data from the Andaman Islands. SuGAR is jointly maintained by the Earth Observatory of Singapore and the Indonesian Institute of Sciences (LIPI). This is Berkeley Seismological Laboratory contribution 2016-5.

Author Contributions Y.H. and R.B. together designed the study and did most of the writing. Y.H. carried out the numerical modelling. P.B., L.F. and E.M.H. collected and processed the daily time series of the SuGAR network. T.I. and T.T. collected and processed the daily time series of the AGNeSS network. K.W. assisted with the modelling strategy. All authors contributed to the interpretations and preparation of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.H. (yhu@seismo.berkeley.edu).

Reviewer Information *Nature* thanks G. Hirth and W. Thatcher for their contribution to the peer review of this work.

METHODS

GNSS data. We collected and processed GNSS time series of 31 SuGAR and 2 AGNeSS stations following the strategy described in Feng *et al.*⁸ using the GPS-Inferred Positioning System and Orbit Analysis Simulation Software (GPSY-OASIS) version 6.2. GNSS daily time series of 11 IGS stations and 3 Memphis stations were downloaded from the Nevada Geodetic Laboratory (Nevada Bureau of Mines and Geology, University of Nevada; <http://geodesy.unr.edu/index.php>, last accessed on 28 July 2015). GNSS daily time series are processed in ITRF2008³⁴.

Over the past two decades a number of large subduction zone earthquakes occurred in Sumatra, including 17 events of moment magnitude $M_w \geq 6.5$ from 2009 up to the IOE (Extended Data Fig. 1). Based on the approach in ref. 9, we take the following steps to derive postseismic displacements from GNSS time series (Extended Data Figs 2 and 3). (1) We correct the time series for the trends of the postseismic transients of the earthquakes before the IOE. We fit the postseismic trends of the previous earthquakes with a logarithmic function of time. (2) We then calculate the long-term secular, annual and semi-annual variations of the time series before the IOE. (3) We correct the post-IOE time series for the trends obtained in step (2). (4) We fit the corrected post-IOE time series using logarithmic and exponential functions of time $a \log(1 + t/\tau_{\log}) + b(1 - \exp(-t/\tau_{\exp}))$, where a and b are constants, t is the time, and τ_{\log} and τ_{\exp} are characteristic time constants of the logarithmic and exponential terms, respectively. τ_{\log} and τ_{\exp} are determined for each GNSS station through a grid search method⁹. (5) We then calculate postseismic displacements between any two time epochs from the fitted postseismic curve (Extended Data Fig. 3). For those stations that were discontinued two or more years after the IOE we calculate the 3-year-postseismic displacements through the extended fitted curve.

We exclude the following ten stations that have data gaps or show patterns of postseismic displacements obviously inconsistent with that of their neighbouring stations (Extended Data Fig. 1). (1) CARI, AITB and NIMT have data gaps of more than 10 days before and after the IOE, 28 January to 23 April 2012, 2–26 April 2012 and 14 March to 26 April 2012, respectively. (2) NGNG and SLBU move westward almost perpendicular to the northward motion of neighbouring stations. PRKB moves southward, opposite to its neighbouring stations. (3) Horizontal displacements at PTLO, TLLU and KTET are more than five times larger than that of neighbouring stations within 100 km. The vertical displacement at BSAT is more than ten times larger than that of nearby stations. The inconsistency in the postseismic deformation pattern of the above stations is probably due to local processes and/or the bias in removing the postseismic trends of local earthquakes before the IOE. The signal-to-noise ratio at the two AGNeSS stations TANG and ACEH increased after 2014 owing to local construction activities. Since our postseismic displacements for TANG and ACEH are calculated through curve fitting based mostly on the time series of 2012–2014, we do not exclude these two AGNeSS stations.

We evaluate test models through calculating the weighted χ^2 misfit:

$$\chi^2 = \frac{1}{N - 1 - \text{d.o.f.}} \sum_{i=1}^N \frac{(G_i - F_i)^2}{\sigma_i^2} \quad (1)$$

where G and F represent GNSS displacement measurements and model predictions, respectively, i represents the station number, the degrees of freedom d.o.f. = 3 in this work are for the three free model parameters, σ_i^2 is the variance of the GNSS observation, and N is the total number of GNSS observations. We use six equally spaced time steps (that is, intervals of 6 months) covering the first three years after the IOE. We calculate the χ^2 misfit of the horizontal and vertical components separately. A linear sum of horizontal and vertical displacements produces preferred models that fit the horizontal components well, but provide a poor fit to the vertical component. Using a higher weight (such as 10) on the vertical component worsens the fit to horizontal components. Therefore we calculate the total effect by a combination of the horizontal components and five times the vertical component.

Finite-element model. The spherical-Earth viscoelastic finite-element model used in this work is based on previous studies of the Chile, Sumatra^{12,13,35,36}, and Cascadia subduction zones¹ and has been reported in refs 13 and 14. The model includes an elastic upper plate, an elastic slab, a viscoelastic mantle wedge, a viscoelastic oceanic asthenosphere and upper mantle (Fig. 2). Cooling and plate models^{37–39} allow for a lithosphere thickness of 50–80 km of the 50–60-million-year-old Indian Ocean plate near the IOE. We thus assume a uniform lithospheric thickness of 50 km, which is also consistent with shear-wave tomography constraints¹⁹ and the depth extent of the coseismic rupture of the IOE^{6,7}. The shear moduli of the elastic lithosphere and viscoelastic upper mantle are assumed to be 48 GPa and 64 GPa, respectively. The Poisson's ratio and rock density are assumed to be 0.25 and $3.3 \times 10^3 \text{ kg m}^{-3}$, respectively, for the entire domain. Viscoelastic relaxation of the upper mantle is represented by the bi-viscous Burgers rheology¹⁵.

On the basis of previous studies¹³ we assume the viscosity of the mantle wedge to be $3 \times 10^{19} \text{ Pa s}$.

The coseismic fault slip of the earthquake derived by Wei *et al.*⁶ is used in this work through the split-node method⁴⁰. Different rupture models^{6,7,16} do not change the fundamental pattern of the predicted co- and postseismic motions at GNSS stations hundreds of kilometres from the rupture area (Extended Data Fig. 4). Except for the top free surface, the other five model boundaries are free in the tangential directions and fixed in the normal direction. Domain boundaries are more than 1,000 km from the rupture zone in the horizontal directions. The bottom of the model is at 660 km depth, approximating the transition zone. The setup of the model boundaries produces negligible numerical artefacts on the deformation of the study area, containing these GNSS stations.

Model tests. We first present explorations of the model space, such as the lithospheric thickness, existence of the slab, and the extent of the oceanic asthenosphere. We examine the contribution of the relaxation in the individual rheological units to the surface deformation. Then we evaluate the potential contributions of afterslip of the fault to the postseismic deformation at GNSS stations. We report the range in three model parameters, the thickness (D_A) and viscosity (η_A) of the oceanic asthenosphere, and the viscosity of the oceanic upper mantle (η_O). Finally we present the temporal change in the postseismic surface deformation in the PM. In the following tests we vary some model parameters and keep other model parameters the same as in the PM, that is, $D_A = 80 \text{ km}$, $\eta_A = 2 \times 10^{18} \text{ Pa s}$, $\eta_O = 10^{20} \text{ Pa s}$, and the viscosity of the mantle wedge $\eta_M = 3 \times 10^{19} \text{ Pa s}$ (Fig. 2). We present model-predicted postseismic displacements at three years after the IOE. Differential surface deformation is calculated by the results of a test model minus that of the PM.

Exploration of the model space. If the oceanic asthenosphere has the same viscosity as the underlying oceanic upper mantle, that is, if we consider models with a homogeneous oceanic upper mantle^{12–14}, a test model with a viscosity of 10^{20} Pa s in the oceanic upper mantle predicts only about half of the observed postseismic horizontal displacements and subsidence of about 2 cm in the forearc area, in the first three years (Extended Data Fig. 5a). Lowering the viscosity (for example, by one order of magnitude; see Extended Data Fig. 5b) improves the fit to the horizontal GNSS data. However, the test model still fails to predict the observed uplift in the forearc region. A weak oceanic asthenosphere is required to produce the observed uplift.

We test a number of model scenarios in which the oceanic asthenosphere is not allowed to extend along the subducting slab, models without a slab, and models with different lithospheric thicknesses. Varying the lithospheric thickness by a couple of tens of kilometres produces negligible changes in the surface deformation (Extended Data Fig. 6a and b). Without the existence of the slab the model predicts additional landward motion near the trench, seaward motion inland, and uplift in the upper plate (Extended Data Fig. 6c). If we assume that the oceanic asthenosphere terminates at the trench and does not extend to greater depths beneath the slab, the differential surface motions three years after the IOE are up to approximately 5 cm near the trench (Extended Data Fig. 6d).

We have constructed test models to study the individual contributions of the rheological units to the surface deformation. We allow viscoelastic relaxation only in one rheological unit using its PM parameter and assume the rest of the domain to be elastic. Although this approach ignores the effects of the viscoelastic flow of other rheological units, it helps to understand the first-order pattern of the deformation that is due to each specific relaxation process.

If we allow viscoelastic relaxation only in the oceanic asthenosphere (Extended Data Fig. 7a), the test model VEA produces horizontal displacements up to more than 50 cm three years after the earthquake. The VEA produces postseismic uplift of more than 7 cm in the northern Sumatra forearc region. If we allow viscoelastic relaxation only in the oceanic upper mantle (Extended Data Fig. 7b), the test model VEO produces up to about 3 cm of the horizontal displacements. The magnitude of the vertical motions in the VEO is smaller than in the VEA, and its direction is opposite to that of the VEA. If we allow viscoelastic relaxation only in the mantle wedge (Extended Data Fig. 7c), the test model VEM produces generally landward motion of less than 5 cm and subsidence of less than 2 cm in the forearc area. Tests on the sensitivity of the surface deformation to variations in the viscosity of the rheological units also indicate that the relaxation in the oceanic asthenosphere has a more important role in controlling the viscoelastic postseismic crustal deformation than that of the underlying upper mantle and the mantle wedge above the subducting slab (results not shown). Note that the IOE induces stresses mostly at shallow depths (for example, less than about 400 km). The PM shows that the three-year-postseismic displacements are up to approximately 2 cm at depths of 400 km, and are negligibly small (less than 1 cm) at greater depths (exceeding 500 km) (results not shown). Therefore, viscoelastic postseismic surface deformation is controlled mainly by relaxation processes in the shallow upper mantle.

We simulate the afterslip after the IOE using a weak shear zone approach¹⁴. In a 2-km-thick shear zone extending down to a depth of 65 km, the maximum depth of the rupture of the IOE⁶, we assume that the locked region is shaped by the 5-m coseismic contour lines within which no afterslip is allowed. Steady-state viscosity η_s in areas outside the locked region is assumed to be 5×10^{17} Pa s (ref. 13). If we do not allow viscoelastic relaxation in the upper mantle (afterslip only), the test model AFS produces substantial horizontal displacements mainly in the vicinities of the rupture area (Extended Data Fig. 7d). The vertical deformation in the AFS is similar to that of the VEO, that is, it produces subsidence in the forearc where postseismic uplift has been observed. If we apply the same weak shear zone to study the IOE-induced afterslip of the megathrust, the resultant change in the surface deformation is no more than 0.4 cm in the three years after the IOE because the stresses on the megathrust induced by the IOE over 200 km away are negligibly small (results not shown).

If we add the contribution from viscoelastic relaxation in the upper mantle using the PM parameters, that is, the model includes the three processes in Extended Data Fig. 7a and c, this afterslip model of $\eta_s = 5 \times 10^{17}$ Pa s produces horizontal displacements at least 50% larger than that in the PM (Extended Data Fig. 8a). Test models with different viscosities in the shear zone produce similar overestimated horizontal GNSS motion (Extended Data Fig. 8b and c). Overestimated motions at GNSS sites are mostly due to afterslip at shallow depths (≤ 50 km) (Extended Data Fig. 8d). Earthquake-induced stress at greater depths (> 50 km) are much smaller, and thus the stress-driven deep afterslip slightly overestimates midfield motions and predicts little changes in the far field (Extended Data Fig. 8e). An afterslip model with a low $\eta_s = 5 \times 10^{17}$ Pa s and a higher η_A (such as $\eta_A = 10^{20}$ Pa s), two orders of magnitude higher than in the PM, produces a better fit to the horizontal GNSS data but worsens the fit to the vertical component (Extended Data Fig. 8f). As afterslip produces subsidence at the northern Sumatra stations, adding its contributions generally increases the model misfits.

In the PM the oceanic asthenosphere extends to greater depths with the downgoing slab. We constructed a test model in which the oceanic asthenospheric layer terminates at the trench⁴¹. Excluding the subducted asthenosphere results in subsidence of up to about 2 cm and southwest seaward displacements of up to about 5 cm in the forearc (Extended Data Fig. 6d). A much lower viscosity (such as $\eta_A = 2 \times 10^{17}$ Pa s; see Extended Data Fig. 9a) or larger thickness (such as $D_A = 200$ km; see Extended Data Fig. 9b) of the asthenosphere is then required to produce a comparable goodness of fit to the land GNSS data.

We assumed a sharp boundary between the lithosphere and the asthenospheric layer and did not include details of the lithosphere–asthenosphere boundary because of the limits of the spatial coverage of the GNSS network. We constructed a test model to study the effect of including a rheological transition between the lithosphere and asthenosphere. In the test model we assume a 20-km-thick transition zone in which the viscosity decreases linearly with depth from 10^{22} Pa s at the bottom of the lithosphere to the preferred 2×10^{18} Pa s of the asthenosphere. Other model parameters are the same as in the PM. This transition-zone model produces a change of no more than 5 cm in surface displacements in areas within 200 km of the rupture area and approximately zero at the land GNSS stations in the first three years after the IOE (Extended Data Fig. 9c). This test thus indicates that the sharpness of the lithosphere and asthenosphere boundary cannot be resolved by the sparse geodetic observations.

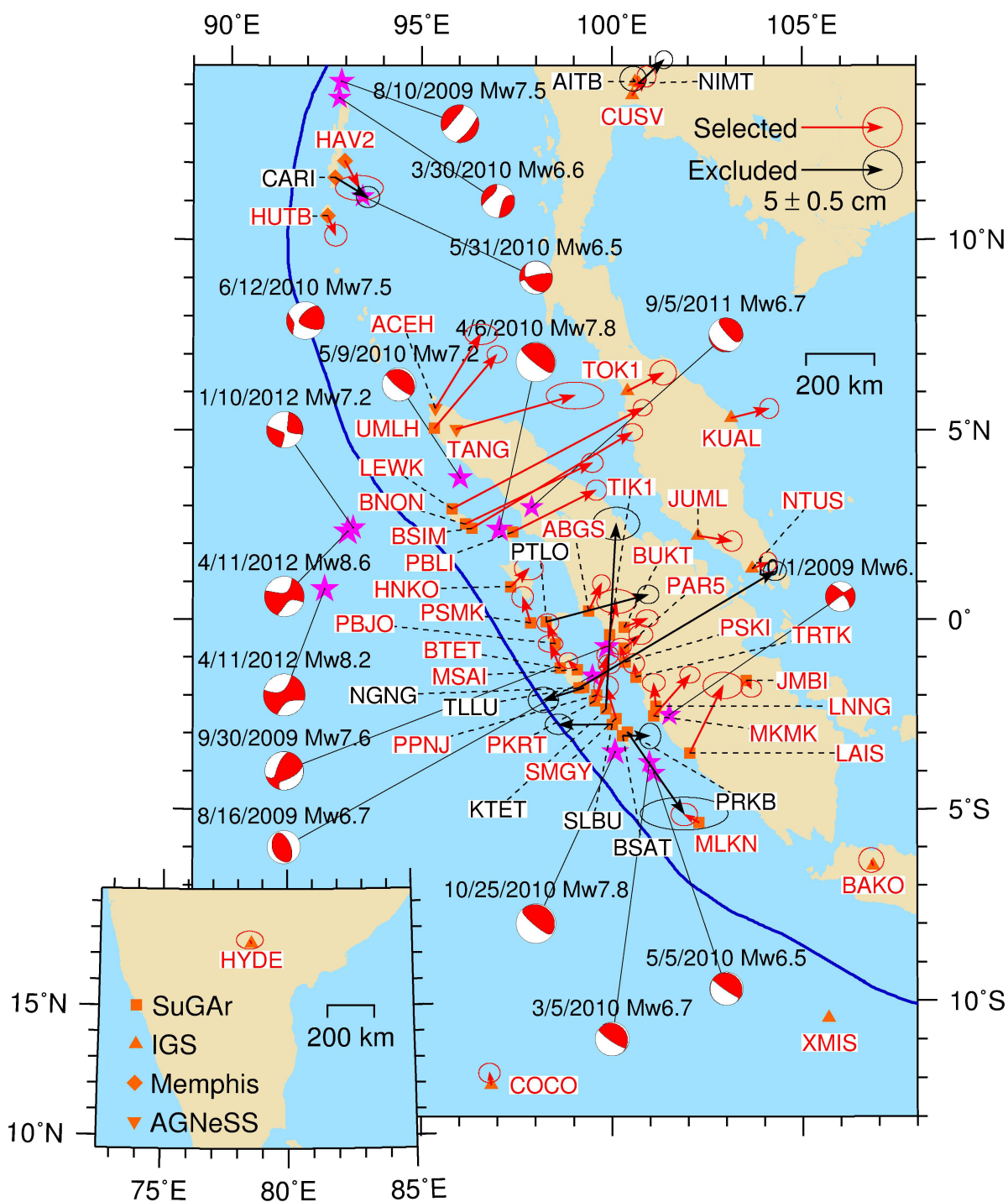
Overall the relaxation in the oceanic asthenosphere is the primary process controlling the postseismic surface deformation and is the only process that produces

the observed uplift in the northern Sumatra forearc. Surface deformation is much more sensitive to the rheological structure below the oceanic lithosphere than to that on the continental side where most of the GNSS stations are located. These test models thus illustrate that the IOE provides a unique opportunity to constrain the rheological structure of the oceanic upper mantle.

Range in model parameters and future predictions in PM. We derive the range of the model parameters by selecting those test models fitting the overall pattern of the GNSS data in both horizontal and vertical directions. The test model best fitting the horizontal GNSS data has $\chi^2 = 5.8$ and does not predict the observed uplift in northwestern Sumatra forearc (Extended Data Fig. 9d). The test model best fitting the vertical GNSS data has $\chi^2 = 6.96$ and overestimates the horizontal data (Extended Data Fig. 9e). We have found that test models with $\chi^2 \leq 5.3$ reproduce the first-order pattern of the GNSS data, that is, misfit of the horizontal components is less than about 20%, and the model predicts more than about 20% of observed uplift at these closest GNSS stations, such as UMLH, LEWK, BNON and BSIM. Test models of $\chi^2 \leq 5.3$ in Fig. 3c thus give the ranges as $D_A = 30$ –200 km, $\eta_A = (0.5$ – $10) \times 10^{18}$ Pa s, and $\eta_O = (0.5$ – $100) \times 10^{20}$ Pa s.

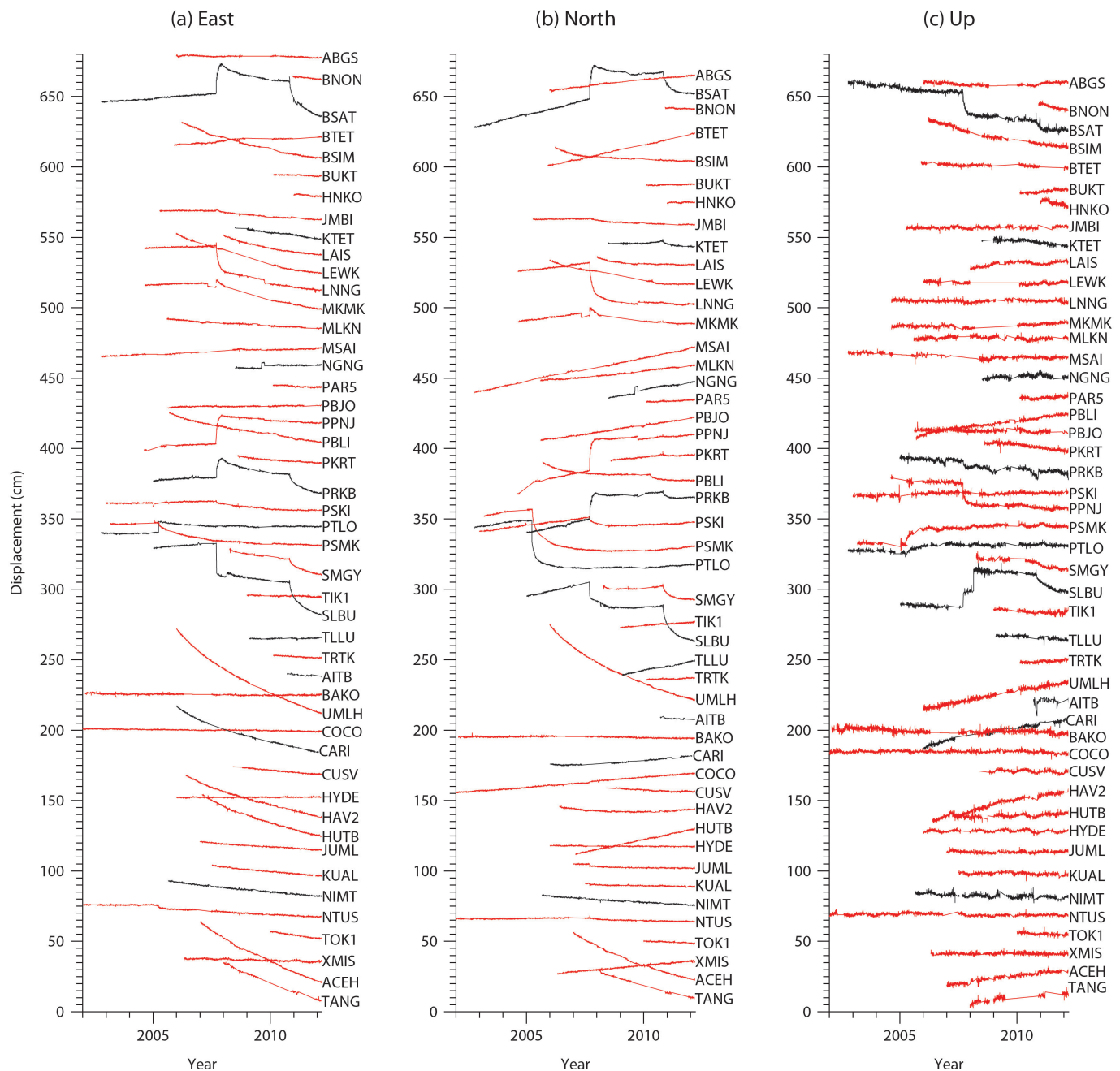
We examine the evolution of the spatial pattern of the predicted viscoelastic postseismic surface deformation in the PM following the IOE (Extended Data Figs 10). The peak horizontal displacements in the upper plate increase from around 10 cm one year after the IOE to more than 50 cm ten years after the IOE (Extended Data Figs 10a–c). Horizontal displacements increase steadily over time and exhibit only small changes in orientation (Extended Data Fig. 10d, e). The vertical surface displacements are generally divided into four uplift–subsidence quadrants, a common pattern of the postseismic deformation following a strike–slip earthquake. An interesting feature is the change in the direction of the vertical displacement in the northeastern quadrant in the continental upper plate (Extended Data Fig. 10a–c, f). In this quadrant the vertical motion one year after the IOE is uplift near the rupture area and subsidence farther inland (Extended Data Fig. 10a, f). The area of the subsidence region shrinks with time, and the uplift region expands.

34. Altamimi, Z., Collilieux, X. & Métivier, L. ITRF2008: An improved solution of the International Terrestrial Reference Frame. *J. Geodyn.* **85**, 457–473 (2011).
35. Hippchen, S. & Hyndman, R. D. Thermal and structural models of the Sumatra subduction zone: implications for the megathrust seismogenic zone. *J. Geophys. Res.* **113**, B12103 (2008).
36. Hu, Y. & Wang, K. Spherical-Earth finite element model of short-term postseismic deformation following the 2004 Sumatra earthquake. *J. Geophys. Res.* **117**, B05404 (2012).
37. Müller, R. D. *et al.* Digital isochrons of the world's ocean floor. *J. Geophys. Res.* **102**, 3211–3214 (1997).
38. Jacob, J., Dymment, J. & Yatheesh, V. Revisiting the structure, age, and evolution of the Wharton Basin to better understand subduction under Indonesia. *J. Geophys. Res. Solid Earth* **119**, 169–190 (2014).
39. Beghein, B., Yuan, Y., Schmerr, N. & Xing, Z. Changes in seismic anisotropy shed light on the nature of the Gutenberg discontinuity. *Science* **343**, 1237–1240 (2014).
40. Melosh, H. J. & Raefsky, A. A simple and efficient method for introducing faults into finite element computations. *Bull. Seismol. Soc. Am.* **71**, 1391–1400 (1981).
41. Hawley, W. B., Allen, R. M. & Richards, M. A. Tomography reveals buoyant asthenosphere accumulating beneath the Juan de Fuca plate. *Science* **353**, 1406–1408 (2016).



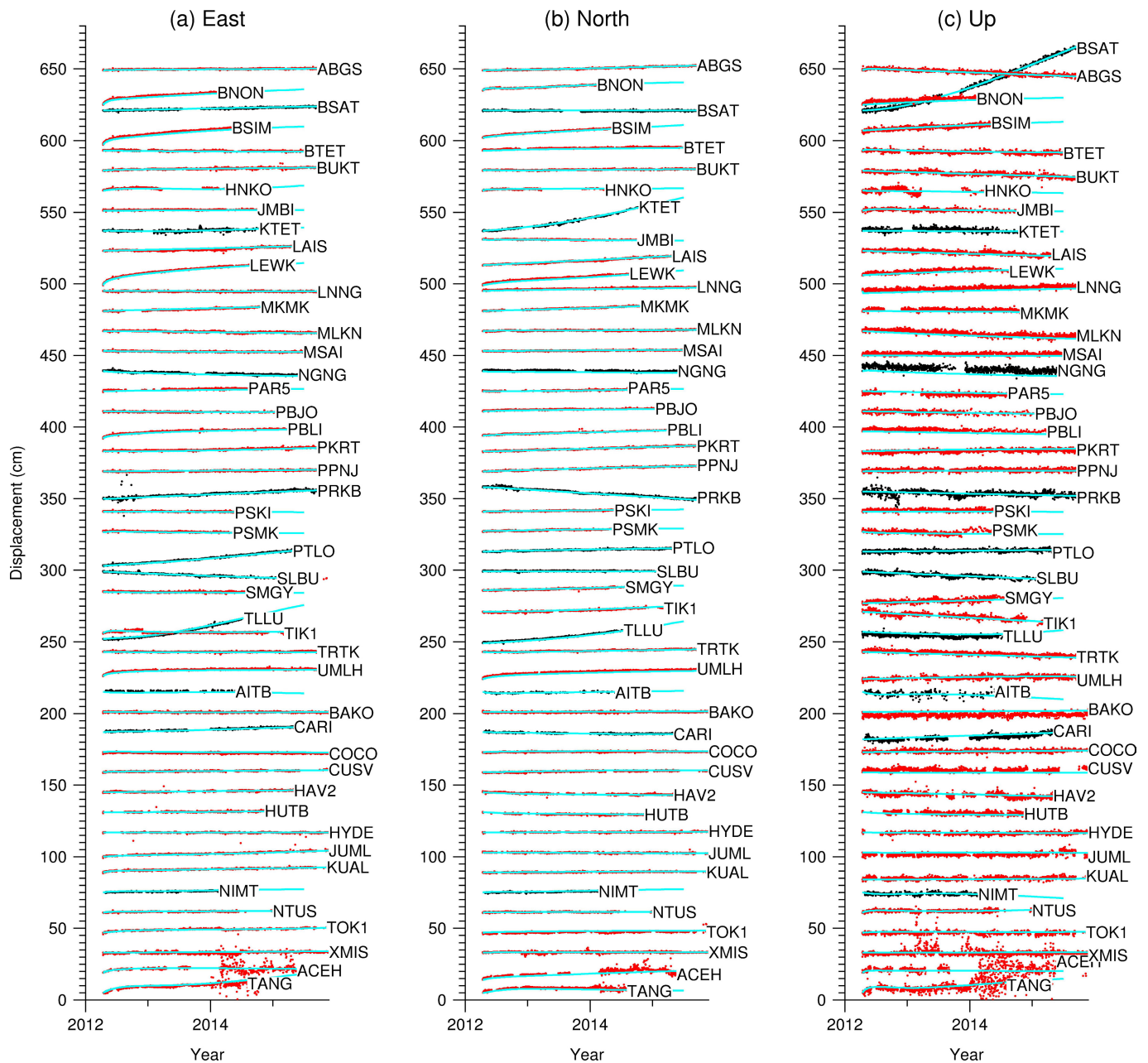
Extended Data Figure 1 | Location of GNSS stations and earthquakes of $M_w \geq 6.5$ from 2009 up to the IOE. Magenta stars represent epicentres of the pre-IOE earthquakes. Red and black arrows represent two-year postseismic displacements at stations that are used or are excluded in this

work, respectively. Station names are labelled with the same colour coding. Solid brown squares, triangles, diamonds and inverted triangles represent GNSS from the SuGAR, IGS, Memphis and AGNeSS networks, respectively.



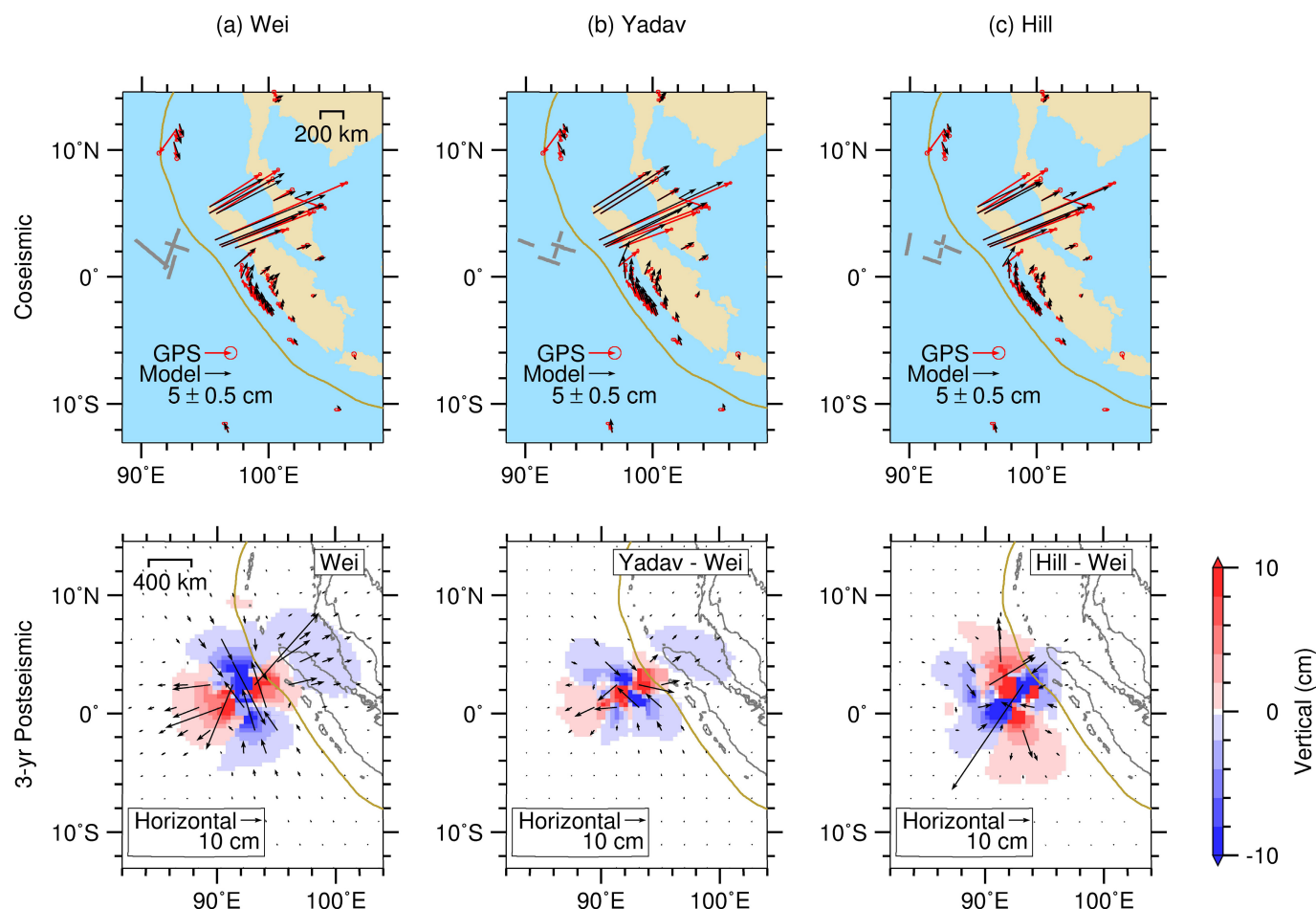
Extended Data Figure 2 | Pre-earthquake daily time series recorded at the GNSS stations shown in Extended Data Fig. 1. a, b and c show the east, north and up components of the time series, respectively. Coseismic static offsets of one day before and after the earthquakes shown in Extended Data Fig. 1 are removed from the time series. The time

series include the total effects of postseismic deformation of previous earthquakes, secular deformation, annual and semi-annual variations. Red and black time series represent those stations that are selected or are excluded in this work, respectively, using the same colour coding as for the station names in Extended Data Fig. 1.



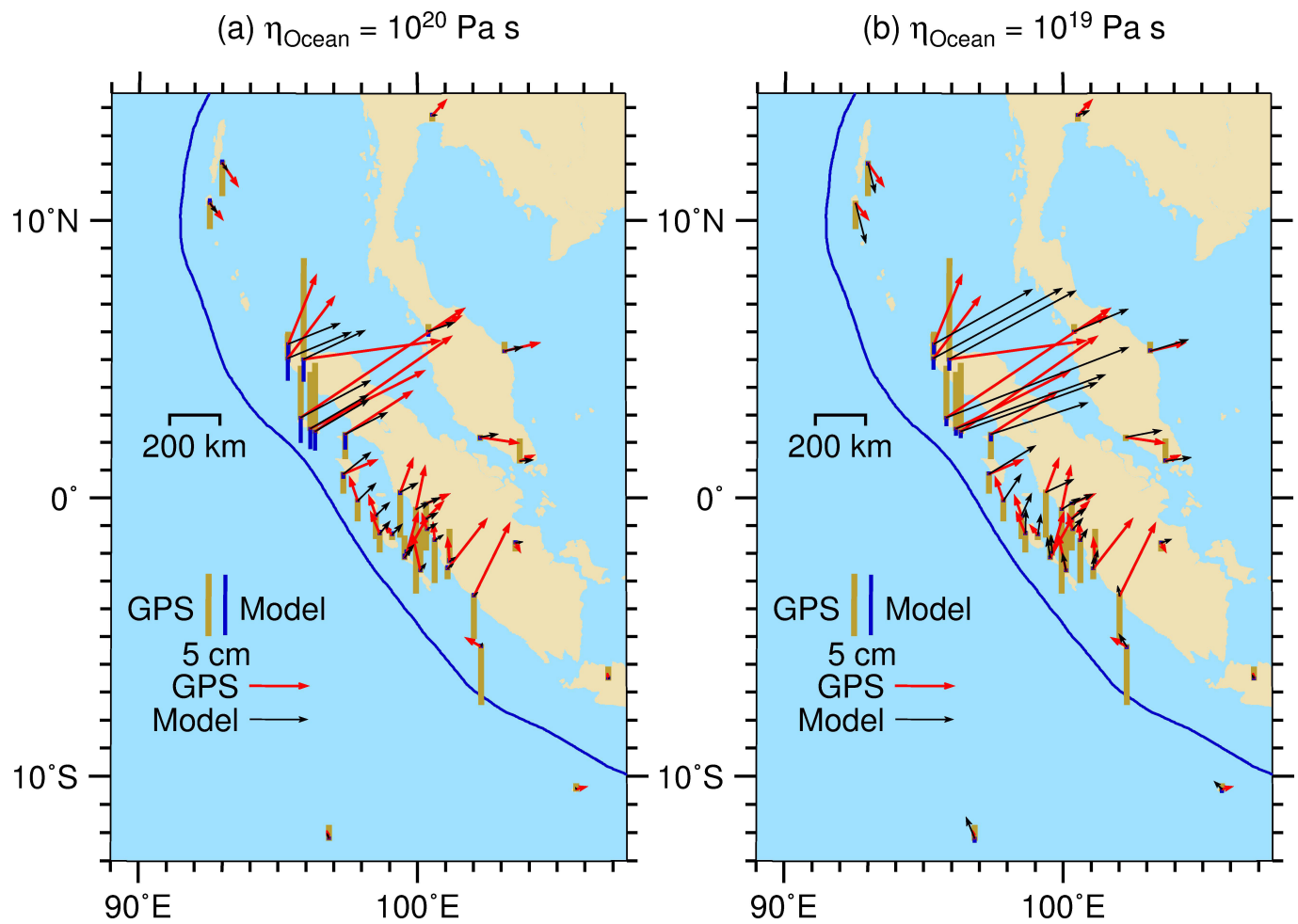
Extended Data Figure 3 | Postseismic GNSS time series after removing postseismic deformation of previous earthquakes, secular motion and seasonal variations. a, b and c show the east, north and up components of the time series, respectively. Red and black lines represent those stations

that are selected or are excluded in this work, respectively. Continuous cyan curves fitted to the postseismic time series are used to constrain our postseismic deformation models.



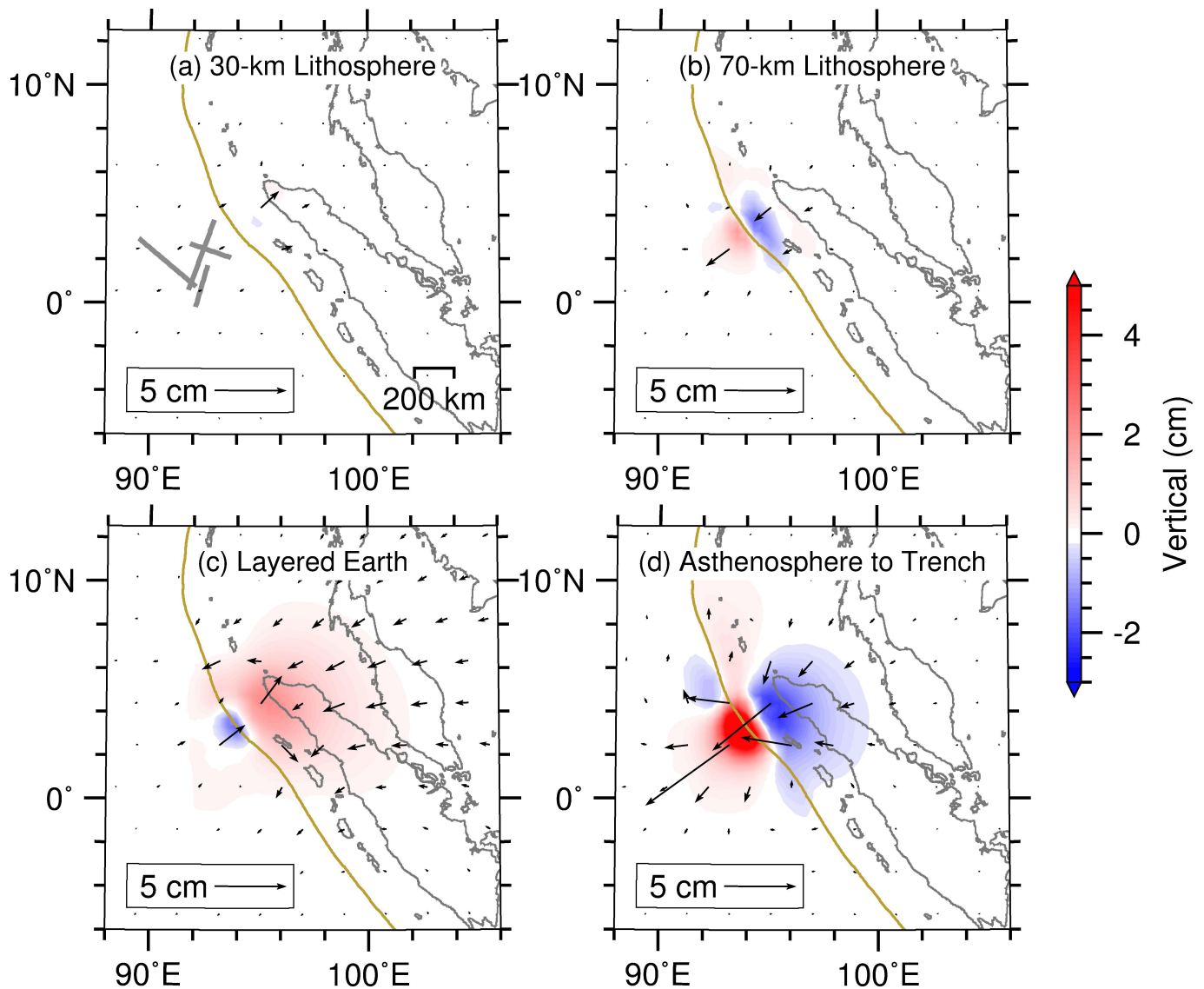
Extended Data Figure 4 | Comparison of different source models of the IOE. **a**, The coseismic slip distribution is from Wei *et al.*⁶, who inverted regional and teleseismic waveform data. Their fault slip model was used in this work. Coseismic GNSS observations are estimated from static offsets of five days before and after the IOE. **b**, The coseismic slip distribution is from Yadav *et al.*¹⁶, who inverted static offsets of 5 days before and after the IOE of daily GNSS data. Model predictions are scaled by 0.8 to fit the coseismic GNSS data better. **c**, The coseismic slip distribution is from Hill *et al.*⁷, who inverted static offsets of about 10 min before and after the IOE of high-rate (one-second rate) GNSS data in the middle field and of

10 days before and after the IOE of daily GNSS data in the far field. Model predictions are scaled by 1.5 to fit the same GNSS data also shown in **a** and **b** better. In the upper panels red and black arrows represent coseismic GNSS observations and model-predicted displacements, respectively. Thick grey lines represent inverted rupture segments of the IOE. In the lower panel of **a** black arrows and colour contours represent model-predicted three-year-postseismic horizontal and vertical displacements, respectively. In the lower panels of **b** and **c** displacements are differenced by the test model minus the model in **a**.



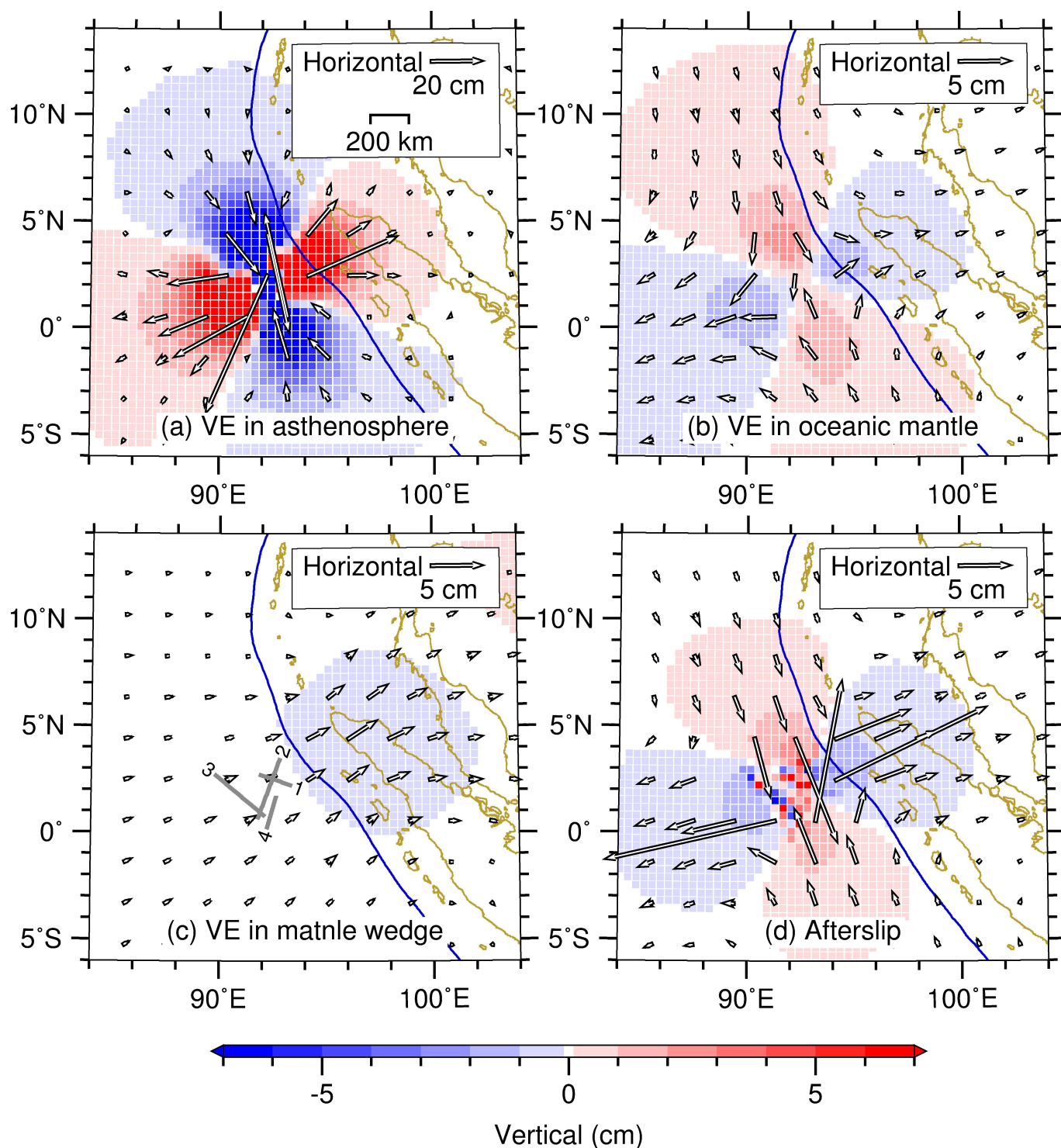
Extended Data Figure 5 | Comparison of three-year-postseismic GNSS observations with predicted displacements in test models of a homogeneous oceanic upper mantle below 50 km without the low-viscosity oceanic asthenosphere. Red and black arrows represent horizontal GNSS observations and horizontal model-predicted

displacements, respectively. Solid brown and blue bars represent vertical GNSS observations and vertical model-predicted displacements, respectively. a, Viscosity of the oceanic upper mantle is 10^{20} Pa s . b, Viscosity of the oceanic upper mantle is 10^{19} Pa s .



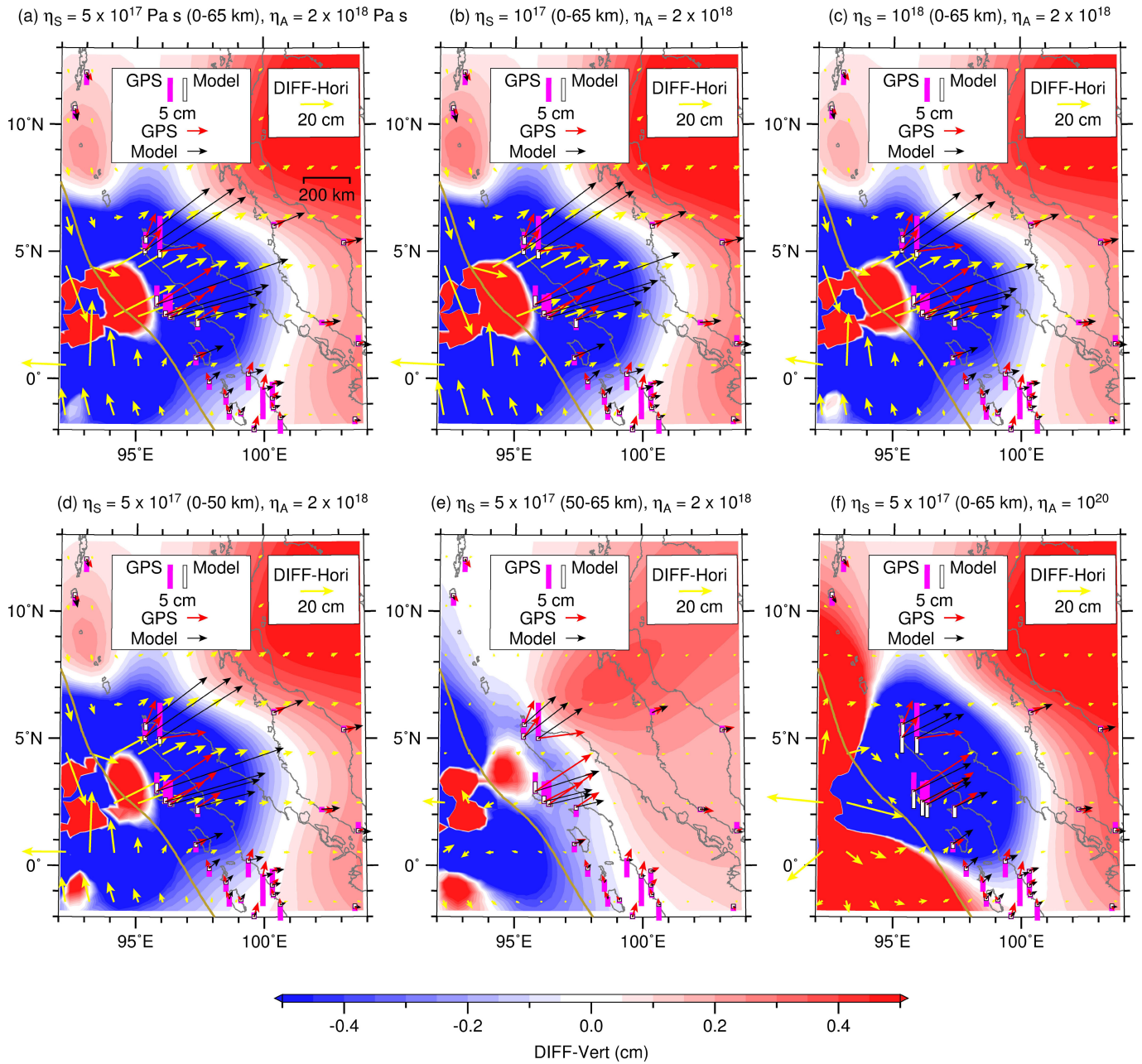
Extended Data Figure 6 | Effects of the extent of the oceanic asthenosphere, layered Earth and variation in the lithospheric thickness on the surface deformation. Displacements are differenced by a test model minus the PM in which the lithospheric thickness, the thickness (D_A) and viscosity (η_A) of the asthenospheric top layer, and the viscosity in the underlying oceanic upper mantle (η_O) are 50 km, 80 km and 2×10^{18} Pa s, and 10^{20} Pa s, respectively. Black and coloured contours represent the horizontal and vertical displacements, respectively.

Thick brown lines outline the location of the trench. **a**, In the test model the lithospheric thickness is assumed to be 30 km, that is, 20 km thinner than in the PM. **b**, Similar to **a** except that the lithospheric thickness is 70 km. **c**, In the test model the slab does not exist. **d**, In the test model the oceanic asthenosphere terminates at the trench and does not extend with the downgoing subducting slab. Thick grey lines in **a** represent rupture segments of the IOE.



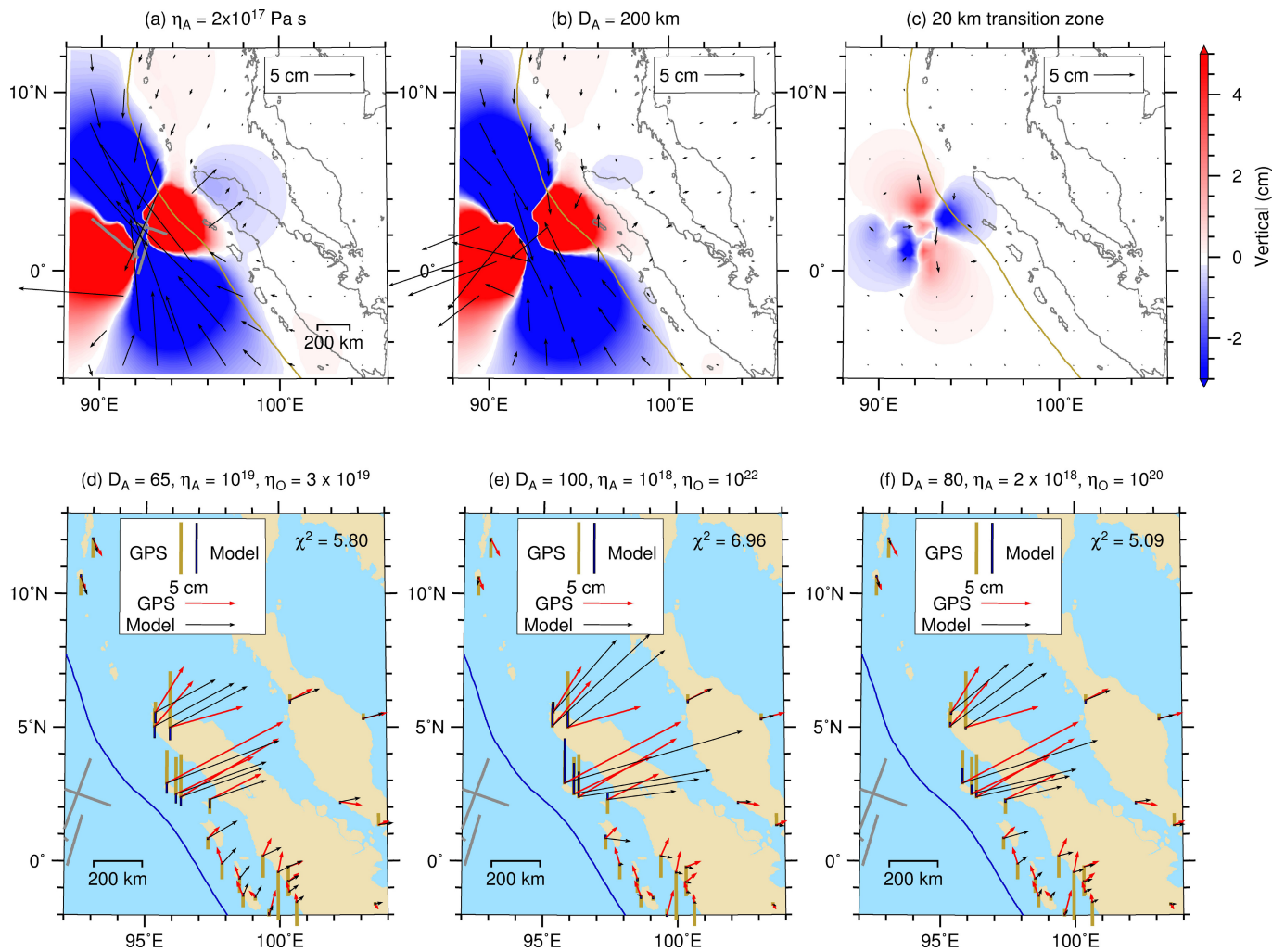
Extended Data Figure 7 | Contributions of viscoelastic relaxation in the rheological units and afterslip of the IOE to the cumulative three-year-postseismic surface deformation. **a**, Surface deformation due to viscoelastic relaxation in the oceanic asthenosphere alone. The continental and oceanic upper mantle are assumed to be elastic. **b**, Surface deformation due to viscoelastic relaxation in the oceanic upper mantle

alone. **c**, Surface deformation due to viscoelastic relaxation in the mantle wedge alone. Thick grey lines represent the rupture segments of the IOE. **d**, Surface deformation due to the modelled afterslip in the shear zone assuming no viscoelastic relaxation elsewhere. Open arrows and colour contours represent horizontal and vertical model-predicted displacements, respectively.



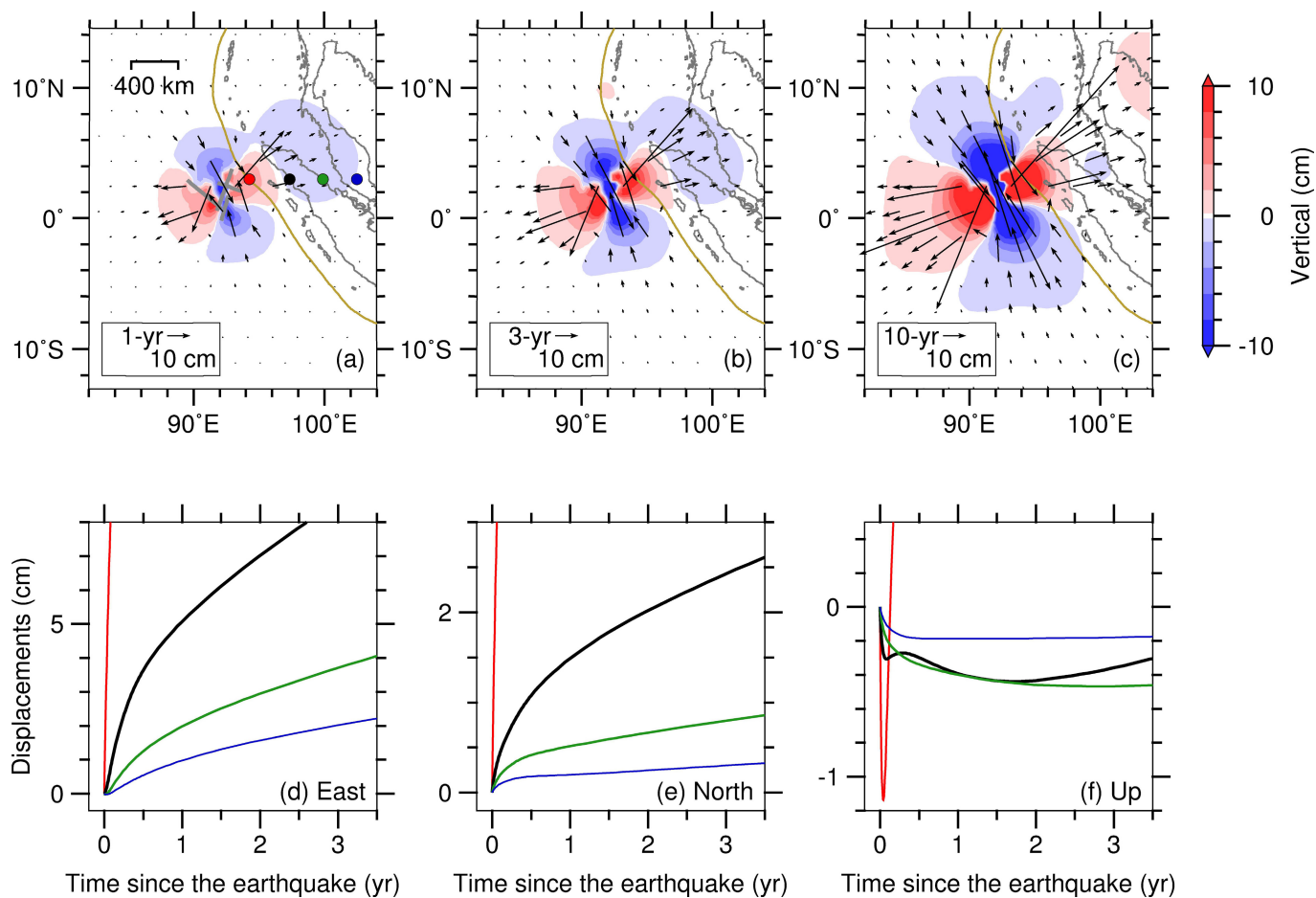
Extended Data Figure 8 | Effects of afterslip after the IOE on the surface deformation. **a**, Steady-state viscosity in the afterslip shear zone η_S is 5×10^{17} Pa s, and $\eta_A = 2 \times 10^{18}$ Pa s. Afterslip is allowed at depths 0–65 km. Red and black arrows represent horizontal GNSS observations and model-predicted displacements, respectively. Solid magenta and white bars represent vertical GNSS observations and model-predicted displacements, respectively. Yellow arrows and colour contours represent differential

horizontal (DIFF – Hori) and vertical (DIFF – Vert) components by the test model minus the PM, respectively. **b**, Similar to **a** except with a low $\eta_S = 10^{17}$ Pa s. **c**, Similar to **a** except with a high $\eta_S = 10^{18}$ Pa s. **d**, Similar to **a** except that the afterslip is allowed only at shallow depths (≤ 50 km) and no deep afterslip. **e**, Similar to **a** except that the afterslip is allowed only at greater depths (50–65 km), and no shallow afterslip. **f**, $\eta_S = 5 \times 10^{17}$ Pa s, and $\eta_A = 10^{20}$ Pa s. Afterslip is allowed at depths of 0–65 km.



Extended Data Figure 9 | Three-year-postseismic displacements due to changes in model parameters and comparison of GNSS observations with predicted displacements. **a**, Surface deformation calculated by the test model minus the PM. In the test model the asthenosphere terminates at the trench and does not extend with the downgoing slab. η_A is one order of magnitude lower than that of the PM. Other model parameters are the same as the PM. Black arrows and contours represent horizontal and vertical three-year-postseismic surface displacements, respectively. Thick grey lines represent rupture segments of the IOE. **b**, Similar to **a** except that η_A is the same as in the PM but $D_A = 200$ km, more than two times thicker than that of the PM. **c**, The sharp boundary between the lithosphere and the asthenosphere in the PM is replaced by a 20-km-thick transition zone in which the viscosity decreases linearly with depth from

10^{22} Pa s at the bottom of the lithosphere to the preferred 2×10^{18} Pa s of the asthenosphere. **d** and **e** are the test models best fitting to the horizontal (Fig. 3a) and vertical (Fig. 3b) GNSS observations, respectively. Red and black arrows represent horizontal GNSS observations and horizontal model-predicted displacements, respectively. Solid brown and blue bars represent vertical GNSS observations and vertical model-predicted displacements, respectively. Thick grey lines represent the rupture segments of the IOE. **f**, Preferred lowest misfit test model (PM) best fitting to both horizontal and vertical GNSS data (Fig. 3c), the same data as in Fig. 5a. Values of the viscosity of the oceanic upper mantle (η_O), thickness (D_A) and viscosity (η_A) of the asthenosphere in each test model are labelled on the top of each plot in **d** and **e**. The value of χ^2 in each test model is labelled as inset text.



Extended Data Figure 10 | Postseismic displacement evolution in the PM. **a**, **b** and **c** show cumulative surface postseismic displacements at one year, three years, and ten years after the earthquake, respectively. Black arrows and contours represent horizontal and vertical displacements, respectively. Thick grey lines in **a** represent the rupture segments of the IOE. **d**, **e** and **f** show the evolution of postseismic displacements in the

east, north and up directions, respectively, at four surface example points of the same latitude 3° N whose locations are at trench (red lines), western (black) and eastern (green) coast of Sumatra, and inland (blue). Locations of these four points are also shown as solid dots in **a** with the same colour coding.

Upper-mantle water stratification inferred from observations of the 2012 Indian Ocean earthquake

Sagar Masuti¹, Sylvain D. Barbot¹, Shun-ichiro Karato², Lujia Feng¹ & Paramesh Banerjee¹

Water, the most abundant volatile in Earth's interior, preserves the young surface of our planet by catalysing mantle convection, lubricating plate tectonics and feeding arc volcanism. Since planetary accretion, water has been exchanged between the hydrosphere and the geosphere, but its depth distribution in the mantle remains elusive. Water drastically reduces the strength of olivine¹ and this effect can be exploited to estimate the water content of olivine from the mechanical response of the asthenosphere to stress perturbations such as the ones following large earthquakes. Here, we exploit the sensitivity to water of the strength of olivine², the weakest and most abundant mineral in the upper mantle, and observations of the exceptionally large (moment magnitude 8.6) 2012 Indian Ocean earthquake³ to constrain the stratification of water content in the upper mantle. Taking into account a wide range of temperature conditions and the transient creep of olivine, we explain the transient deformation in the aftermath of the earthquake that was recorded by continuous geodetic stations along Sumatra as the result of water- and stress-activated creep of olivine. This implies a minimum water content of about 0.01 per cent by weight—or 1,600 H atoms per million Si atoms—in the asthenosphere (the part of the upper mantle below the lithosphere). The earthquake ruptured conjugate faults down to great depths⁴, compatible with dry olivine in the oceanic lithosphere. We attribute the steep rheological contrast to dehydration across the lithosphere–asthenosphere boundary, presumably by buoyant melt migration to form the oceanic crust.

Water is heterogeneously distributed in Earth's interior among nominally anhydrous minerals (for example, olivine and pyroxene), melts, and fluids. The principal source of water in the upper mantle is from its dissolution in olivine and other minerals as a point defect. Olivine, the most abundant upper-mantle mineral, constitutes a water reservoir that may exceed the mass of the oceans⁵.

Since the 2012 moment magnitude (M_w) 8.6 Indian Ocean earthquake ruptured the entire oceanic lithosphere^{3,4,6}, the postseismic deformation (Fig. 1) provides strong constraints on the rheological properties of the oceanic lithosphere–asthenosphere system. This is in contrast to most postseismic deformation, which occurs in the continental regions, and for which the influence of the lower crust needs to be taken into account in order to infer the rheological properties of the upper mantle⁷.

Several mechanisms may explain the postseismic deformation, including accelerated viscoelastic flow in the asthenosphere and triggered slip (afterslip) along the Indian Ocean coseismic faults or the Sunda megathrust. Vertical deformation is sensitive to the gradient of the horizontal displacements, providing useful constraints with which to identify the likely source of deformation^{8–10}. Afterslip around the coseismic slip area creates horizontal motion compatible with observations, but predicts subsidence of the forearc islands, opposite to what is observed (Fig. 1, Extended Data Figs 1 and 4). Afterslip on the Sunda megathrust creates the correct polarity of vertical motion but the thrust

motion gives the reverse sense of horizontal displacements. Viscoelastic flow confined in a finite-thickness asthenosphere predicts the correct polarity of vertical displacements and the correct azimuth of horizontal displacements. Probably both afterslip and viscous flow contribute to the displacements and their mechanical coupling must be included in the analysis.

The plastic deformation of olivine is accommodated by a combination of diffusion creep and dislocation creep^{2,7,11} and the steady-state rheology (the constitutive stress–strain rate relationship) of olivine is a thermally activated flow law of the form^{11,12}

$$\dot{\epsilon} = A\sigma^n d^{-m} (C_{\text{OH}})^r \exp\left(-\frac{Q+pV}{RT}\right)$$

where $\dot{\epsilon}$ is the strain rate, A is a pre-exponential factor, σ is the deviatoric stress, n is the stress exponent, Q is the activation energy, R is the universal gas constant, T is the temperature, V is the activation volume, p is the confining pressure, d is grain size, m is the grain size exponent, and C_{OH} and r are the water concentration and its exponent. Diffusion creep is associated with $n = 1$ and $m = 2–3$, and dislocation creep with $n = 3–5$ and $m = 0$. The temperature weakening and pressure hardening in the Arrhenius law define the top and bottom boundaries of the mechanical asthenosphere, respectively. The strain rates from the two mechanisms add up, but one of them dominates depending on stress, temperature, pressure and grain size^{12,7}. Ubiquitous observations of seismic anisotropy in the upper mantle^{13,14} suggest that dislocation creep is the dominant deformation mechanism at these depths, which occurs for grain sizes of $d = 6$ mm (assuming an activation energy of 280 kJ mol^{-1} for diffusion creep) and above (Fig. 2b), compatible with observations of grain size ranging from 5 mm to 20 mm in mantle xenoliths². Because of the competition between grain size reduction by dislocation creep and grain growth during diffusion creep the two creep mechanisms may coexist at steady state², but the dominance of dislocation creep is enhanced during the postseismic period because of the large stress perturbation from the mainshock (Fig. 2b, Extended Data Fig. 3).

Laboratory experiments show that olivine creep exhibits a rapid transient before reaching steady-state creep^{15–17} and previous studies have suggested that postseismic signals cannot be explained with realistic Earth properties without including the transient effect^{18–20}. The underlying mechanism for the transient behaviour is not fully understood but some experiments suggest that it may be caused by the transition in the mechanism of strain accommodation in dislocation creep¹⁷ or by the evolution of internal stress distribution in diffusion creep^{21,22}. Here, we propose an addition to the rheology of olivine that includes these transient effects in the diffusion creep or dislocation creep regime (see Methods)

$$\dot{\epsilon}_K = A_K(\sigma - 2G_K\epsilon_K)^n (C_{\text{OH}})^r d^{-m} \exp\left(-\frac{Q+pV}{RT}\right)$$

¹Earth Observatory of Singapore, Asian School of the Environment, Nanyang Technological University, Singapore. ²Department of Geology and Geophysics, Yale University, New Haven, Connecticut, USA.

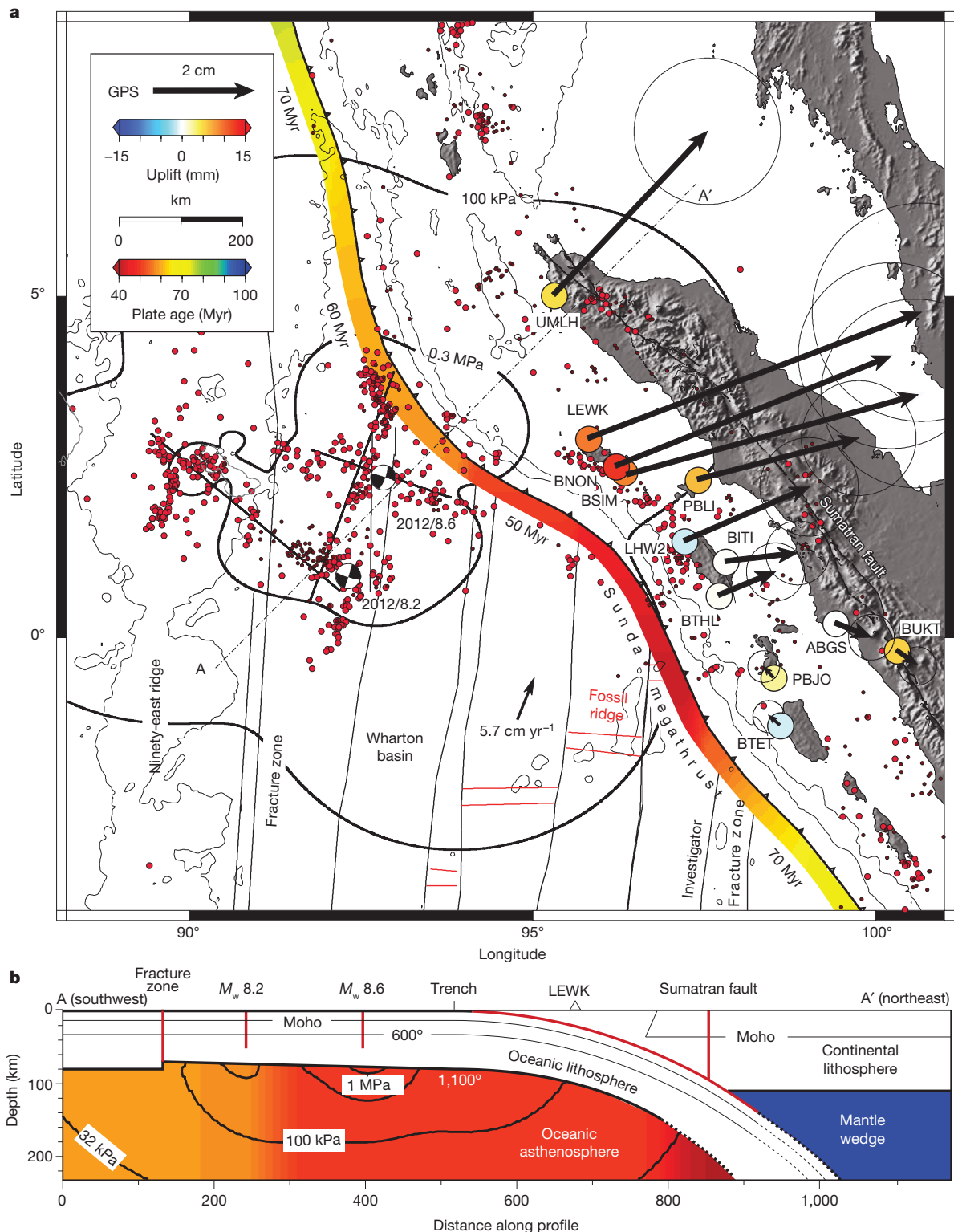


Figure 1 | Postseismic deformation and aftershocks (red dots, NEIC catalogue) 1 year after the 2012 M_w 8.6 Indian Ocean earthquake. **a, The stations of the Sumatran GPS Array north of the Equator moved landwards while the others moved seawards (black vectors with 2σ uncertainties). The northern stations overall uplifted (coloured circles, warm colours for uplift). Stations LEWK, BNON and BSIM on a forearc island produced the most uplift. Stations that subsided are shown as cool-coloured circles. The source mechanisms of the mainshock and largest aftershock are from**

where $\dot{\epsilon}_K$ is the transient strain and G_K is a work hardening coefficient. The formulation is compatible with the linear Burgers rheology that combines a Kelvin and a Maxwell element in series for $n = 1$, which

the Global Centroid Moment Tensor Catalog (black beachballs). The plate age²⁴, which controls the temperature at depth, is shown along the Sunda trench. The coseismic stress change driving the postseismic relaxation is shown in contours of 300 kPa and 30 kPa at 100 km depth (thick black profile). Thin black lines represent the bathymetry. **b**, Cross-section along the A–A' profile shown in **a** of the plate age, coseismic stress change and slab geometry. The red segments indicate faults in the brittle layer. Moho, Mohorovičić discontinuity.

corresponds to the transient of diffusion creep. (Here, the subscript K corresponds to the constitutive properties of the Kelvin element.) For larger power exponents and $m = 0$, the formulation represents the

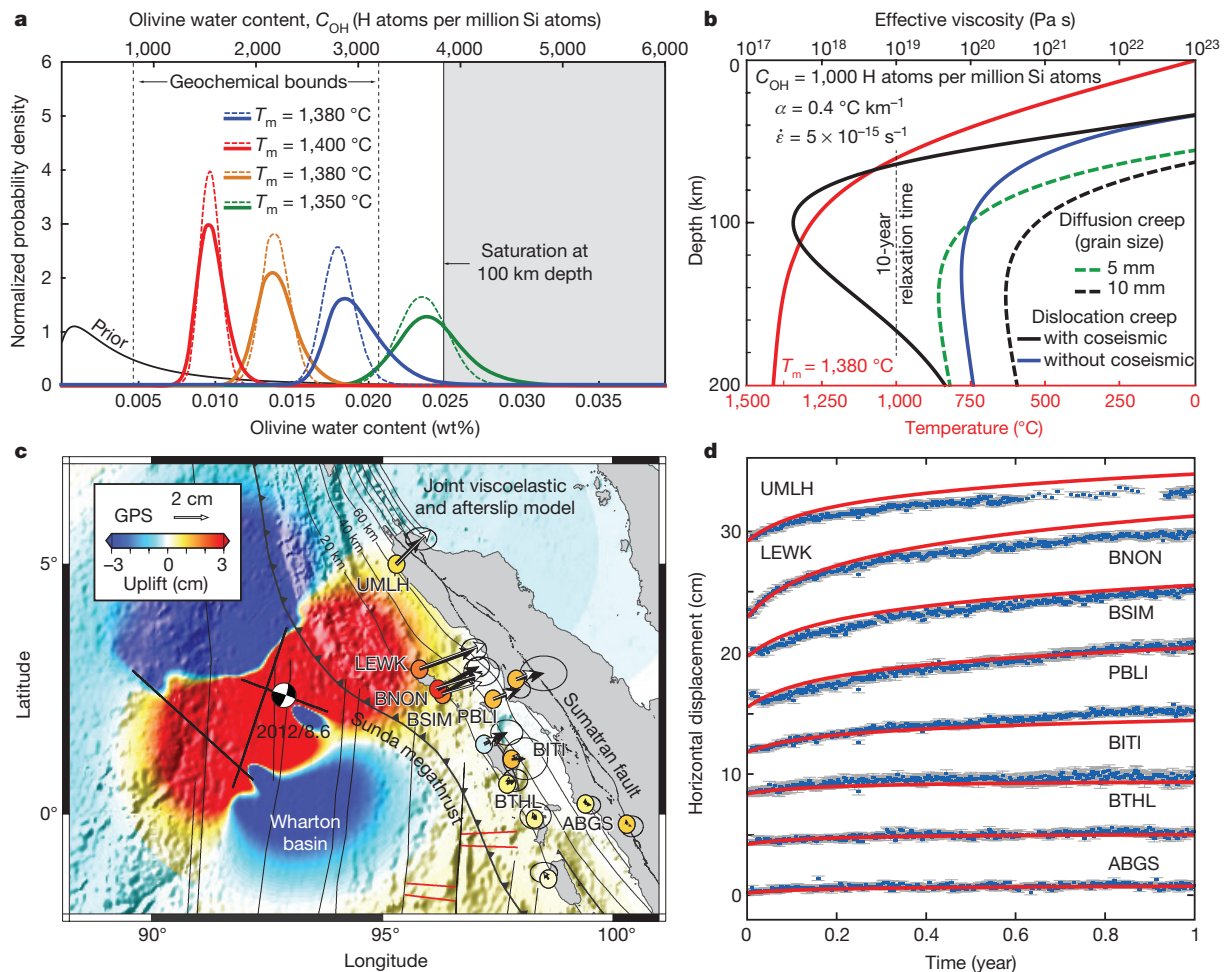


Figure 2 | Probability density of water content in asthenosphere olivine from modelling of postseismic deformation. **a**, Marginal (solid profile) and conditional (dashed profile) probability density of water content for different mantle temperatures T_m , with (red, yellow, green) and without (blue) transient creep. The prior density of water content from laboratory experiments (solid black profile) is a log-normal distribution with a mean value of 600 H atoms per million Si atoms and a standard deviation of 1. Geochemical estimates (vertical dashed profiles) of water content in the MORB source provide independent bounds. **b**, Effective viscosity for the dislocation (solid profiles) and diffusion creep (dashed profiles) regimes. The temperature (red profile) is for a cooling half space with

adiabatic gradient $\alpha = 0.4^\circ\text{C km}^{-1}$ for a 60-million-year-old plate. The rheological parameters are from ref. 12. The strength of dislocation creep is dynamically reduced owing to coseismic stress changes from the mainshock (black solid profile). **c**, Observed (as in Fig. 1) and modelled (white arrows for horizontal and background colour for vertical) displacements after one year. The slab contours are from the USGS Slab1.0. **d**, Time series of GPS (blue squares with $\pm 1\sigma$ error bars) and modelled (red profile) displacements. The fit to other stations of the Sumatran Global Positioning System (GPS) Array (SuGAR) network is shown in Extended Data Fig. 7.

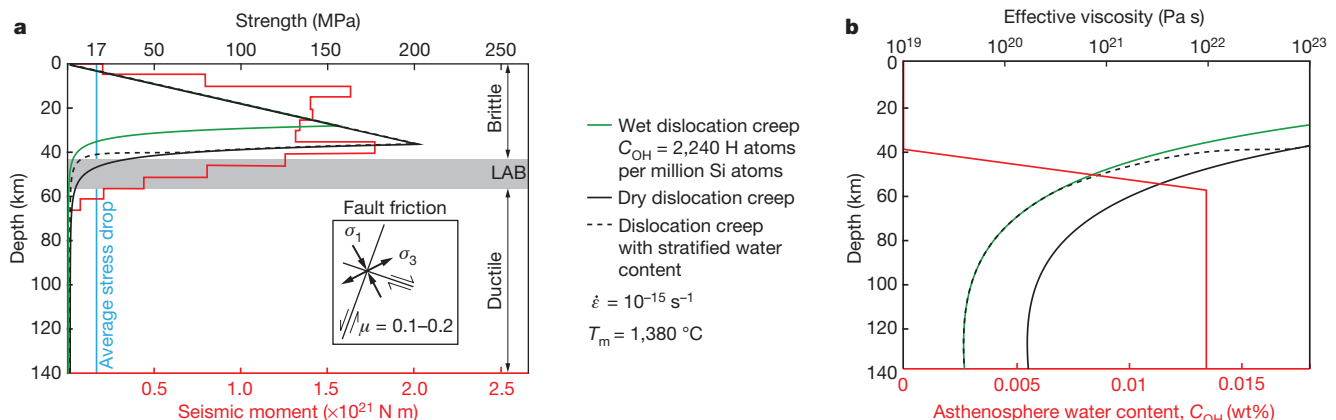


Figure 3 | Strength of the lithosphere and earthquake rupture. **a**, Moment release from the mainshock from ref. 4 (red profile), average stress drop (blue profile) and lithosphere strength assuming wet olivine (green profile), dry olivine (black solid profile) or stratified water content (black dashed profile). The brittle strength assumes a low effective coefficient of friction

based on the near-orthogonal conjugate faults that ruptured during the earthquake (inset). **b**, Effective viscosity for wet olivine, dry olivine and stratified water content (profiles as in **a**) using the constitutive equation of ref. 12. The water content (red profile) changes across the lithosphere–asthenosphere boundary (LAB) between depths of about 45 km and 60 km.

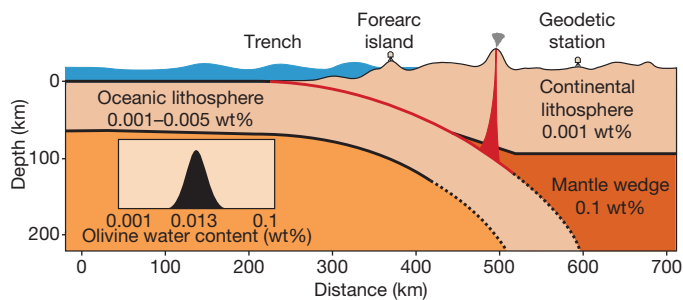


Figure 4 | Schematic of water content distribution across a subduction zone. The oceanic brittle lithosphere is dry (0.001–0.005 wt%) and the oceanic asthenosphere is wet (about 0.013 wt%), with a water content given by the probability density (inset) assuming the mantle temperature $T_m = 1,380^\circ\text{C}$. The water content in the continental lithosphere and the mantle wedge are from ref. 30.

transient of dislocation creep; both are thermally activated (parameters such as A , r , Q and V for transient creep may be different from those for steady-state creep, but for simplicity, we assume that these are the same between transient and steady-state creep except for the pre-exponential factors A and A_K). The transient and steady-state strain rates add up to relax the same stress, leading to a more rapid initial deformation than with steady-state creep alone (Extended Data Fig. 2). At steady state the internal stress is relaxed ($\sigma - G_K \epsilon_K = 0$), so the transient flow does not affect the long-term strength of the upper mantle.

We propose a physical model for the geodetic observations where viscoelastic relaxation in the asthenosphere and afterslip on the Indian Ocean coseismic faults work in concert to relax the coseismic stress perturbation from the mainshock and the largest aftershock²³. We build a three-dimensional model with a brittle lithosphere that subducts below the Sunda trench (Fig. 2c) and rides above a low-viscosity asthenosphere (see Methods). The strength of the upper mantle is controlled by friction in the lithosphere and the temperature and pressure dependence of olivine flow in the asthenosphere (we use a cooling half-space and the oceanic plate age model of ref. 24). We choose the effective friction coefficient (Fig. 3a) to obtain a brittle-to-ductile transition depth of 45 km, based on the deep rupture of the mainshock^{4,6}. We model afterslip with rate-strengthening friction^{25,26} down to depths of 70 km, below and around fault areas of negative coseismic stress change.

We use a Bayesian approach to explore the likelihood of the physical parameters that control olivine flow and afterslip by assimilating the constraints from our geodetic observations. We test a range of water content in olivine from 0.0003 wt% to 0.04 wt% (50–6,000 H atoms per million Si atoms), mantle temperatures from $1,350^\circ\text{C}$ to $1,400^\circ\text{C}$ with an adiabatic temperature gradient of $0.4^\circ\text{C km}^{-1}$ (Fig. 2b, see Methods), and reference velocities for the reference rate of afterslip in the range $V_0 = 0\text{--}2.75 \mu\text{m s}^{-1}$. We choose grain size so that diffusion and dislocation creep add up to the ambient interseismic strain rate at steady state. We first evaluate the bivariate probability density and then integrate out the afterslip parameter to produce the single-variate probability density for water content in olivine (Fig. 2a). The result incorporates the uncertainties originating from both sources of deformation, that is, postseismic slip and viscous flow after the earthquake. We extend the parameter search manually as necessary to investigate the role of the background strain rate, the mantle temperature and the transient parameters of olivine (see Methods).

We can explain the geodetic observations with the laboratory-derived constitutive properties of olivine and realistic *in situ* conditions with a water content in olivine ranging from 0.01 wt% (1,600 H atoms per million Si atoms) for a mantle temperature of $1,400^\circ\text{C}$ to 0.023 wt% (3,660 H atoms per million Si atoms) for a mantle temperature of $1,350^\circ\text{C}$ (Fig. 2, Extended Data Figs 6 and 7). The best-fit model for $T_m = 1,380^\circ\text{C}$ has $A = 10^{0.56} \text{ s}^{-1} \text{ MPa}^{-n} \text{ m}^{-m}$, $Q = 418.5 \text{ kJ mol}^{-1}$, $V = 11 \text{ cm}^3 \text{ mol}^{-1}$, $r = 1.2$, $m = 0$ and $n = 3.5$ for dislocation creep¹², and $A = 10^6 \text{ s}^{-1} \text{ MPa}^{-n} \text{ m}^{-m}$, $Q = 280 \text{ kJ mol}^{-1}$, $V = 4 \text{ cm}^3 \text{ mol}^{-1}$, $r = 1$,

$m = 3$, $d = 6 \text{ mm}$ and $n = 1$ for diffusion creep¹¹, and water content of $\text{COH} = 2,240$ H atoms per million Si atoms in equation (1) and the same parameters except $A_K = A/2$ and $G_K = G$ in equation (2). There is little tradeoff between water content and afterslip speed (Extended Data Fig. 6) and the maximum likelihood value of the afterslip parameter falls in the range found in other tectonic settings²⁶. Considering the large uncertainties in the water content of mid-oceanic ridge basalt (MORB) source and the uncertainties of water partition coefficients in the mantle rock assemblage⁵, we find our estimated water content in olivine is in good agreement with geochemical estimates for the MORB source⁵ in the range 0.005–0.02 wt% (800–3,200 H atoms per million Si atoms). Ignoring the transient creep of olivine introduces bias in estimates of water content (Fig. 2a). With mantle temperatures below $1,350^\circ\text{C}$, our estimates of water content are above full water saturation at 100 km depth (Fig. 2a).

The geodetic data are also sensitive to the thickness of the asthenosphere, which is controlled by the pressure hardening of olivine through the activation volume or the transition to a high-pressure phase at 410 km. Models with an activation volume lower than $5 \text{ cm}^3 \text{ mol}^{-1}$ for dislocation creep fail to explain the observations, indicating a finite thickness of the mechanical asthenosphere during the postseismic transient. As diffusion creep has a low activation volume^{11,12}, the result is a further indication that dislocation creep is the dominant deformation mechanism in the asthenosphere during the early postseismic transient. The geodetic time series are better modelled assuming a low ambient stress, indicating that diffusion creep and dislocation creep have similar strength at steady state (see Methods), compatible with grain sizes of 6–10 mm depending on the background strain rates assumed (Fig. 2b).

The 2012 M_w 8.6 Indian Ocean earthquake propagated into the lithosphere–asthenosphere boundary with an average stress drop of 17 MPa (ref. 6), and a slip distribution tapering off near the brittle-to-ductile transition, at 45–65 km depth⁴ (Fig. 3a). The rupture took place on near-perpendicular conjugate faults^{3,4,6}, indicating a low effective coefficient of friction (see Methods). The thick lithosphere is compatible with the strength of dry olivine (Fig. 3). These observations imply that water content is stratified with a steep gradient at the lithosphere–asthenosphere boundary (Fig. 3b). Dehydration of the lithosphere may happen by buoyant partial melt to form the oceanic crust, as the solubility of volatiles is higher in the liquid phase^{1,27}.

Partial melting is often invoked to explain weak rheology and high electrical conductivity^{28,29}. However, the deformation occurring during and after the 2012 M_w 8.6 Indian Ocean earthquake can be explained by a steep rheological gradient across the lithosphere–asthenosphere boundary, controlled by temperature and water stratification without weakening caused by partial melting (Fig. 4). Application of the proposed methodology to other great and giant earthquakes can refine our estimates of water content in other tectonic regions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 April; accepted 16 August 2016.

Published online 10 October 2016.

- Karato, S.-I., Paterson, M. S. & Fitzgerald, J. D. Rheology of synthetic olivine aggregates—influence of grain size and water. *J. Geophys. Res.* **91**, 8151–8176 (1986).
- Karato, S. & Wu, P. Rheology of the upper mantle: a synthesis. *Science* **260**, 771–778 (1993).
- Yue, H., Lay, T. & Koper, K. D. En echelon and orthogonal fault ruptures of the 11 April 2012 great intraplate earthquakes. *Nature* **490**, 245–249 (2012).
- Wei, S., Helmberger, D. & Avouac, J.-P. Modeling the 2012 Wharton basin earthquakes off Sumatra: complete lithospheric failure. *J. Geophys. Res.* **118**, 3592–3609 (2013).
- Hirschmann, M. M. Water, melting, and the deep Earth H_2O cycle. *Annu. Rev. Earth Planet. Sci.* **34**, 629–653 (2006).
- Hill, E. *et al.* The 2012 M_w 8.6 Wharton Basin sequence: a cascade of great earthquakes generated by near-orthogonal, young, oceanic mantle faults. *J. Geophys. Res.* **120**, 3723–3747 (2015).

7. Bürgmann, R. & Dresen, G. Rheology of the lower crust and upper mantle: evidence from rock mechanics, geodesy, and field observations. *Annu. Rev. Earth Planet. Sci.* **36**, 531–567 (2008).
8. Bruhat, L., Barbot, S. & Avouac, J. P. Evidence for postseismic deformation of the lower crust following the 2004 M_w 6.0 Parkfield earthquake. *J. Geophys. Res.* **116**, B08401 (2011).
9. Rousset, B., Barbot, S., Avouac, J. P. & Hsu, Y.-J. Postseismic deformation following the 1999 Chi-Chi earthquake, Taiwan: implication for lower-crust rheology. *J. Geophys. Res.* **117**, B12405 (2012).
10. Rollins, J. C., Barbot, S. & Avouac, J.-P. Mechanisms of postseismic deformation following the 2010 El Mayor-Cucapah earthquake. *Pure Appl. Geophys.* **54**, 1305–1358 (2015).
11. Hirth, G. & Kohlstedt, D. L. in *Inside the Subduction Factory* (ed. Eiler, J.), Vol. 138, 83–105 (Geophysical Monographs, AGS, 2003).
12. Karato, S. & Jung, H. Effects of pressure on high temperature dislocation creep in olivine. *Phil. Mag.* **83**, 401–414 (2003).
13. Gaherty, J. B., Jordan, T. H. & Gee, L. S. Seismic structure of the upper mantle in a central pacific corridor. *J. Geophys. Res.* **101**, 22291–22309 (1996).
14. Montagner, J.-P. & Tanimoto, T. Global anisotropy in the upper mantle inferred from the regionalization of phase velocities. *J. Geophys. Res.* **95**, 4797–4819 (1990).
15. Hanson, D. R. & Spetzler, H. A. Transient creep in natural and synthetic, iron-bearing olivine single crystals: mechanical results and dislocation microstructures. *Tectonophysics* **235**, 293–315 (1994).
16. Chopra, P. High-temperature transient creep in olivine rocks. *Tectonophysics* **279**, 93–111 (1997).
17. Karato, S.-I. in *Ice Age Geodynamics: A New Perspective* (ed. Wu, P.) 351–364 (Trans Tech Publications, 1998).
18. Freed, A. M., Herring, T. & Bürgmann, R. Steady-state laboratory flow laws alone fail to explain postseismic observations. *Earth Planet. Sci. Lett.* **300**, 1–10 (2010).
19. Pollitz, F. F. Transient rheology of the uppermost mantle beneath the Mojave desert, California. *Earth Planet. Sci. Lett.* **215**, 89–104 (2003).
20. Wang, K., Hu, Y. & He, J. Deformation cycles of subduction earthquakes in a viscoelastic earth. *Nature* **484**, 327–332 (2012).
21. Raj, R. Transient behavior of diffusion-induced creep and creep rupture. *Metall. Trans. A* **6**, 1499–1509 (1975).
22. Mackwell, S. J., Kohlstedt, D. L. & Paterson, M. S. The role of water in the deformation of olivine single crystals. *J. Geophys. Res.* **90**, 11319–11333 (1985).
23. Barbot, S. & Fialko, Y. A unified continuum representation of postseismic relaxation mechanisms: semi-analytic models of afterslip, poroelastic rebound and viscoelastic flow. *Geophys. J. Int.* **182**, 1124–1140 (2010).
24. Müller, R., Sdrolias, M., Gaina, C. & Roest, W. Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochem. Geophys. Geosyst.* **9**, Q04006 (2008).
25. Barbot, S., Fialko, Y. & Bock, Y. Postseismic deformation due to the M_w 6.0 2004 Parkfield earthquake: stress-driven creep on a fault with spatially variable rate-and-state friction parameters. *J. Geophys. Res.* **114**, B07405 (2009).
26. Marone, C., Scholz, C. H. & Bilham, R. On the mechanics of earthquake afterslip. *J. Geophys. Res.* **96**, 8441–8452 (1991).
27. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
28. Karato, S. Does partial melting reduce the creep strength of the upper mantle? *Nature* **319**, 309–310 (1986).
29. Naif, S., Key, K., Constable, S. & Evans, R. Melt-rich channel observed at the lithosphere–asthenosphere boundary. *Nature* **495**, 356–359 (2013).
30. Karato, S. Rheology of the earth's mantle: a historical review. *Gondwana Res.* **18**, 17–45 (2010).

Acknowledgements We are grateful to our LIPI collaborators who maintain the SuGAR network, including J. Encillo, I. Suprihanto, D. Prayudi and B. Suwargadi. Raw SuGAR data are available for download at <ftp://eos.ntu.edu.sg/SugarData>. We thank M. Sambridge for sharing his Neighborhood Algorithm software. The modelling software used in this study is hosted at www.geodynamics.org/cig/software/relax with support from the Computational Infrastructure for Geodynamics. This research was supported by the National Research Foundation of Singapore under the NRF Fellowship scheme (National Research Fellow Awards numbers NRF-NRFF2013-04 and NRF-NRFF2010-064) and by the Earth Observatory of Singapore, the National Research Foundation, and the Singapore Ministry of Education under the Research Centres of Excellence initiative. This is EOS publication 120.

Author Contributions S.M., S.D.B. and S.-i.K. conducted the study and wrote the manuscript. L.F. prepared the GPS time series. P.B. develops the SuGAR network.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.D.B. (sbarbot@ntu.edu.sg).

Reviewer Information *Nature* thanks G. Hirth and W. Thatcher for their contribution to the peer review of this work.

METHODS

GPS data processing. We process the SuGAR raw data using the GPS-Inferred Positioning System and Orbit Analysis Simulation Software (GIPSY-OASIS) software (hosted at <https://gipsy-oasis.jpl.nasa.gov/>). To extract the postseismic transient due to the 2012 M_w 8.6 Indian Ocean earthquake, we identify and remove the signals from other earthquakes in the entire available time series for each component of each station. In this procedure, we simultaneously estimate the linear long-term rates, coseismic, postseismic, and seasonal signals by nonlinear least-squares fitting. A complete description of these steps can be found in refs 6 and 31. We isolate the postseismic deformation from the 2012 Indian Ocean earthquake sequence by removing the contribution of all other identified sources. The resulting time series is shown in Extended Data Fig. 1.

Formulation of the transient rheology of olivine. The effect of the transient creep of rocks, whether for diffusion creep or dislocation creep, is well known in the laboratory and some studies suggest that it is important to incorporate it in models of postseismic deformation^{18–20} or postglacial rebound³². Several attempts have been made to formulate a rheological law for transient creep^{33–37}, but they have important shortcomings. For example, laws such as^{33,34}

$$\sigma = A \varepsilon^q \dot{\varepsilon}^r \quad (1)$$

where A , q and r are positive constitutive parameters, or^{35,36}

$$\sigma = B \left[1 - \exp \left(- \left(\frac{\varepsilon}{\varepsilon_0} \right)^s \right) \right] \dot{\varepsilon}^r \quad (2)$$

where B , ε_0 , r and s are positive constants, can create only a single transient creep episode, as plastic strain systematically increases, and therefore they do not describe transient creep for repeated earthquake stress perturbations. Some other laws, such as the so-called Andrade creep³⁷

$$\dot{\varepsilon} = \frac{ABnt^{n-1}}{1 + At^n} + \dot{\varepsilon}_{ss} \quad (3)$$

where $\dot{\varepsilon}_{ss}$ is the steady-state strain rate and A and B are positive parameters, are singular³⁸ at $t = 0$, for $n > 1$. More fundamentally, laws such as equation (3) fail to satisfy fundamental principles such as time-frame invariance (the same flow should be predicted for any clock).

There are several ways to formulate a rheology for the transient creep of olivine by invoking work hardening. For simplicity, we propose a generalization of the Burgers rheology. The Burgers rheology may appropriately represent transient creep in the diffusion creep regime, and we adapt it to be compatible with the power-law steady state of dislocation creep.

In a Burgers material, where the Kelvin element and the Maxwell element are in series, the inelastic strain rate can be written

$$\dot{\varepsilon} = \dot{\varepsilon}_K + \dot{\varepsilon}_M \quad (4)$$

where $\dot{\varepsilon}_K$ is the inelastic strain rate in the Kelvin element and $\dot{\varepsilon}_M$ is the inelastic strain rate in the Maxwell element. The strain rate of the Maxwell element is given by

$$\dot{\varepsilon}_M = A(\sigma)^n (C_{OH})^r d^{-m} \exp \left(- \frac{Q + pV}{RT} \right) \quad (5)$$

where the parameters are defined in the main text. In general, we can then formulate a generalized rheology for the Kelvin element of the form

$$\dot{\varepsilon}_K = f_K(\sigma - G_K \varepsilon_K) \quad (6)$$

where σ is the deviatoric stress (the same as in the Maxwell element in series) and $(\sigma - G_K \varepsilon_K)$ is the stress in the Kelvin dashpot. Because the functional form f_K is unknown, we assume for simplicity that it is the same as for the steady-state creep

$$\dot{\varepsilon}_K = A_K(\sigma - G_K \varepsilon_K)^n (C_{OH})^r d^{-m} \exp \left(- \frac{Q + pV}{RT} \right) \quad (7)$$

where G_K is a work hardening coefficient. In the absence of detailed laboratory data on wet olivine, we assume that the thermodynamic parameters for transient and steady-state creep are the same (for wet olivine, ref. 16 shows that flow law parameters are similar between transient and steady-state creep except for the pre-exponential factor). However, laboratory experiments indicate that if transient creep is due to the transition of slip from the soft slip system to the hard slip system of olivine, transient creep should be less sensitive to the water content²².

At the background stress σ_0 , if the flow is at steady state (that is, $\dot{\varepsilon}_K = 0$), we have $\sigma_0 - G_K \varepsilon_K^0 = 0$, where ε_K^0 is the cumulative strain at the background stress. After a stress perturbation from an earthquake, the stress changes from σ_0 to $\sigma_0 + \Delta\sigma$ and the rate of transient creep instantaneously becomes $\dot{\varepsilon}_K = f_K(\sigma_0 + \Delta\sigma - G_K \varepsilon_K^0) = f_K(\Delta\sigma)$. So transient creep, unlike steady-state creep, is not sensitive to the background stress during the postseismic transient, unless it was not at steady state. Multiple stress perturbations also lead to multiple transient creep episodes.

We explore the predictions of the proposed rheology in a spring-slider system. Together with the constitutive relationship, conservation of momentum leads to the system of coupled ordinary differential equations

$$\dot{\sigma} = -G(\dot{\varepsilon}_M + \dot{\varepsilon}_K)$$

$$\dot{\varepsilon}_M = A(\sigma)^n (C_{OH})^r d^{-m} \exp \left(- \frac{Q + pV}{RT} \right) \quad (8)$$

$$\dot{\varepsilon}_K = A_K(\sigma - G_K \varepsilon_K)^n (C_{OH})^r d^{-m} \exp \left(- \frac{Q + pV}{RT} \right)$$

We solve these equations numerically with unit stress perturbations and typical laboratory-derived constitutive properties (Extended Data Fig. 2). The transient creep accelerates the immediate relaxation that follows the stress perturbation and the subsequent relaxation is slower, compared to when operating at steady state only. The hardening coefficient G_K controls the amount of strain that is relaxed by transient creep.

We generalize the constitutive relationship for transient creep for three-dimensional deformation with isotropic rheology. The constitutive relationship becomes

$$(\dot{\varepsilon}_K)_{ij} = A_K(q)^{n-1} (C_{OH})^r d^{-m} \exp \left(- \frac{Q + pV}{RT} \right) Q_{ij} \quad (9)$$

where the subscripts i and j are the tensor indices

$$Q_{ij} = \sigma_{ij} - 2G_K(\varepsilon_K)_{ij} \quad (10)$$

is the internal deviatoric stress tensor and $q^2 = Q_{kl}Q_{kl}$ is the norm of the internal deviatoric stress tensor (we use Einstein's summation convention). We implement these constitutive relationships in the community code Relax (www.geodynamics.org/cig/software/relax), such that we can simulate three-dimensional models of postseismic deformation that includes afterslip, viscoelastic flow and the transient creep of olivine in a self-consistent manner. Extended Data Figure 2b and c shows the predicted time series at a few SuGAR stations for various values of G_K .

Bounds on pre-stress. The dynamics of postseismic deformation is sensitive to the stress preceding the coseismic perturbation in the asthenosphere because of the power-law dependence in the stress–strain rate relationship. Our mechanical model for the power-law flow of olivine includes (1) the background strain rate due to the shear arising from the vertical gradient of horizontal flow, which is associated with the long-term oblique subduction of the oceanic lithosphere below Sumatra, and (2) the pure shear associated with internal deformation by conjugate strike-slip faulting of the Wharton basin along the diffuse boundary between the Indian and Australian plates. We convert the background strain rate to pre-stress, accounting for the water content, temperature and the other physical parameters of olivine rheology. We assume a homogeneous velocity gradient from the surface (plate velocity of 5.6 cm yr^{−1} oriented 10° N; Fig. 1) to a relative fixed transition zone. The orientation of the conjugate faults in the Wharton basin suggests that the deformation in the lithosphere is horizontal pure shear³⁹ with principal stress orientation between −30° N and −25° N for the compressive component and between 60° N and 65° N for the extensive component (Fig. 3).

The long-term strain rate is the sum of the linear (diffusion creep) and nonlinear (dislocation creep) strain rate contributions

$$\dot{\varepsilon} = \dot{\varepsilon}_{\text{diffusion}} + \dot{\varepsilon}_{\text{dislocation}} \quad \text{or} \quad \dot{\varepsilon} = \tau \left(\frac{1}{\eta_{\text{diffusion}}} + \frac{1}{\eta_{\text{dislocation}}} \right) \quad (11)$$

where $\eta_{\text{diffusion}}$ and $\eta_{\text{dislocation}}$ are the effective viscosities for diffusion and dislocation creep, respectively. While the linear viscosity due to diffusion creep has little effect on the initial postseismic deformation and cannot be directly estimated (Fig. 2b), it does affect the background stress

$$\tau = \frac{\eta_{\text{dislocation}}}{1 + \frac{\eta_{\text{dislocation}}}{\eta_{\text{diffusion}}}} \dot{\varepsilon} \quad (12)$$

So a low-viscosity diffusion creep reduces the background stress and the associated strain rate for dislocation creep.

We consider two end-member models for the background strain rate $\dot{\epsilon}$. In a first model, plate motion is accommodated in the mantle across a 100-km-thick region, leading to a background strain rate of $2 \times 10^{-14} \text{ s}^{-1}$. In a second model, the region is 400 km thick, leading to a background strain rate of $5 \times 10^{-15} \text{ s}^{-1}$. Considering these models one at a time, we investigate a range of strain rates for dislocation creep ranging from full background strain rate (corresponding to vanishing diffusion creep in the model) to $1 \times 10^{-17} \text{ s}^{-1}$ (corresponding to dominant diffusion creep at steady state). We find that our geodetic data are best explained with a dislocation strain rate of the order of 10^{-17} s^{-1} to 10^{-16} s^{-1} , corresponding to grain size between 6 mm and 10 mm depending on the background strain rate. This indicates that diffusion creep and dislocation creep have similar strengths in the asthenosphere at steady state. The strength of diffusion creep is lower than the one for dislocation creep for grain sizes lower than 6 mm (Fig. 2b).

The recent great and giant earthquakes along the Sunda megathrust affected the background stress preceding the 2012 earthquake sequence⁴⁰. To estimate this effect, we evaluate the deviatoric stress change in the asthenosphere at the time of the 2012 event caused by the nearby 2004 M_w 9.2 Aceh–Andaman and the 2005 M_w 8.6 Nias earthquakes (Extended Data Fig. 3). The deviatoric stress near the rupture area of the 2012 event at 100 km depth due to these earthquakes is about two orders of magnitude smaller than the coseismic stress change. For simplicity we ignore the stress caused by the Sunda megathrust earthquakes in our simulations. **Lithosphere strength.** The strength of the brittle lithosphere can be evaluated from the orientation of the conjugate faults using Byerlee's law ($\tau = \mu' \sigma_n$). The effective coefficient of friction is given by

$$\mu' = \frac{1}{\tan 2\theta} \quad (13)$$

where θ is the angle between the fault and the principal stress. The direction of the principle stress should bisect the conjugate faults. In the Indian Ocean earthquake rupture, the main conjugate faults are nearly orthogonal (Figs 1 and 3), providing an estimate of the effective coefficient of friction in the range 0.1–0.2. The exact orientation of the rupture fault is still subject to debate⁶, affecting our estimate of the effective friction coefficient. The inferred small friction coefficient in the Wharton basin reduces the strength of the brittle lithosphere (Fig. 3a).

The thickness of the lithosphere can be independently estimated from the depth of the coseismic rupture of the 2012 M_w 8.6 Indian Ocean earthquake as substantial coseismic slip occurred down to 60 km depth⁴. The reason for the great depth extent of the rupture is unclear, but strong dynamic weakening from frictional melting may have allowed the rupture to penetrate below the seismogenic zone. The depth range from 45 km to 60 km where coseismic slip tapers from its maximum value may correspond to the lithosphere–asthenosphere boundary (Fig. 3a). **Forward models of postseismic deformation.** We build a three-dimensional model of the lithosphere and asthenosphere in which the flow parameters depend on depth, except in the subducted slab where viscosity is infinite, resulting in a three-dimensional rheological model. The background depth-dependent viscosity is thermally activated following a half-space cooling model

$$T(z, t) = T_0 + T_m \operatorname{erf} \left(\frac{z}{\sqrt{4Kt}} \right) + \alpha(z - z_0)H(z - z_0) \quad (14)$$

with $T_0 = 0^\circ \text{C}$, where T_m is the mantle temperature (a free parameter), the plate age t varies spatially following the model of ref. 24, the thermal diffusivity is $K = k\rho_m^{-1}C_p^{-1}$ with conductivity $k = 3.138 \text{ W m}^{-1} \text{ K}^{-1}$, the specific heat is $C_p = 1.171 \text{ kJ kg}^{-1} \text{ K}^{-1}$, the density is $\rho_m = 3.330 \text{ kg m}^{-3}$, $\alpha = 0.4^\circ \text{C km}^{-1}$ is the adiabatic temperature gradient, $H(z)$ is the Heaviside function, and $z_0 = 100 \text{ km}$ is the depth of the thermal boundary layer for a 60-million-old oceanic plate (Fig. 2b). In our models, most of the postseismic deformation occurs around 100 km depth, where the coseismic stress change is high, so the adiabatic temperature gradient has almost no effect on our water content estimates. The elastic slab is constructed using the Slab 1.0 model⁴¹ with a uniform thickness of 80 km. Because the accelerated flow is concentrated in areas of high coseismic stress change, immediately below the mainshock, the effect of the slab does not greatly affect the fit to the geodetic data.

We create models of postseismic relaxation where the coseismic stresses from the mainshock and the largest aftershock are potentially relaxed by both afterslip in the lithosphere and viscoelastic flow in the asthenosphere, depending on the rheological parameters. We use the coseismic slip models of ref. 4 to produce the stress perturbation and to define the geometry of faults for afterslip. In our relaxation models, both afterslip and viscoelastic deformation are driven by stress and the coupling between the two mechanisms is taken into account. Afterslip is allowed only on the faults that ruptured coseismically, around the areas of negative

coseismic stress change. We assume uniform friction properties outside the rupture area, as only near-field data would motivate finer-tuned models.

We model afterslip with rate-strengthening friction, a simplification of the rate-and-state friction law that is adequate to represent triggered aseismic slip at steady state^{8–10}. In the rate-strengthening approximation the afterslip velocity V is given by²⁵

$$V = 2V_0 \sinh \frac{\Delta\tau}{(a-b)\sigma} \quad (15)$$

where $\Delta\tau$ is the stress change after the mainshock, $(a-b)$ is the steady-state friction parameter and σ is the effective normal stress on the fault. We assume a uniform $(a-b)\sigma = 6 \text{ MPa}$ in our simulations and we explore values of V_0 that best explain the geodetic observations in combination with other mechanisms of deformation (Extended Data Fig. 6).

The calculations are performed using the Relax software (www.geodynamics.org), which employs a spectral method to evaluate the quasi-static deformation caused by stress perturbations. The method incorporates an equivalent body-force representation of dislocations that allows us to include detailed finite slip distributions. The equivalent body-force method is advantageous compared to other approaches such as finite-element methods because meshing of the domain around three-dimensional faults is automatic. The numerical approach has been validated using comparisons with analytic solutions for fault slip and comparisons with finite-element solutions in simple geometric settings^{23,25,42,43}. Another advantage of the Relax software and the equivalent body-force method is the ability to simulate several physical mechanisms of deformation simultaneously and to include nonlinear rheologies for plastic flow^{8–10}.

Limits of single-mechanism deformation models. Afterslip as a single mechanism of postseismic deformation predicts horizontal deformation compatible with the observation. However, the vertical displacement predicted by afterslip in the nearest stations is opposite to observations, suggesting another active process of postseismic deformation, such as viscoelastic relaxation in the asthenosphere (Extended Data Fig. 4).

Viscoelastic flow in the upper mantle can occur by diffusion creep, dislocation creep, or a combination of both⁷. The physical properties of dislocation creep in olivine are well known from a wealth of laboratory experiments documenting the effect of temperature, pressure⁷ and water⁴⁴. The properties of diffusion creep of olivine are less well known because of its great sensitivity to grain size. Several field samples show that both diffusion creep and dislocation creep act in tandem to deform mylonite shear zones below the brittle-to-ductile transition⁴⁵. Experimental work suggests that dislocation creep of olivine is the dominant mechanism of deformation in the upper-mantle conditions^{2,46}. But competition between grain growth during diffusion creep and grain size reduction by dislocation creep can promote comparable strain rate at these depths at equilibrium. In this case, dislocation creep should dominate postseismic deformation following a large stress perturbation because of the power-law stress–strain rate dependence. As the stress is reduced by postseismic relaxation the contributions of both mechanisms should become similar again.

Following these considerations we first tested the potential of dislocation creep to explain the postseismic transient. Initially, we ignored the transient creep of olivine. We find that nonlinear viscoelastic deformation produces vertical displacements compatible with observations and horizontal displacements in the correct azimuth. However, the amplitude of horizontal displacements is lower than observed. We conclude that steady-state dislocation creep cannot explain the entire set of observations. Instead, we find that only a combination of afterslip on the Indian Ocean coseismic faults and viscoelastic flow can satisfactorily explain the GPS observations.

Inverse models of postseismic deformation. We use a Bayesian approach to estimate the water content in olivine, assimilating prior information from geochemical estimates and additional constraints from geodetic observations. The posterior probability density is⁴⁷

$$\sigma_M(\mathbf{m}) = \nu \exp \left[-\frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^T \mathbf{C}_D^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) \right] \rho_M(\mathbf{m}) \quad (16)$$

where ν is a constant, \mathbf{d}_{obs} is the GPS data vector, $\mathbf{g}(\mathbf{m})$ is the forward model based on parameters \mathbf{m} , \mathbf{C}_D is the data covariance matrix and ρ_M is the prior information. Geochemistry provides limits on the water content in the mid-ocean ridge basalt source. We include these constraints on water content using a log-normal distribution (Fig. 2a). For the afterslip parameter V_0 , we assume the uniform prior $1/V_0$, which corresponds to the limit of a log-normal distribution for infinite variance (Extended Data Fig. 6a). We assume that the geodetic data are independently and normally distributed. In principle, the joint probability density of all physical parameters can be estimated, but we limit our exploration to two physical parameters

at a time to reduce the computational burden. In a first step, we jointly estimate the water content in olivine and the rate-strengthening parameters for afterslip, and assume all other parameters to be fixed. In a second step, we change the *in situ* parameters manually to explore a range of conditions.

We sample the posterior probability density using the Neighbourhood algorithm⁴⁸, a derivative-free Markov chain Monte Carlo method. The inversion tool, dubbed Relax-Miracle, uses Relax for the forward models. We benchmark the approach using a synthetic data set using the GPS network configuration of the Sumatra GPS Array. We create a forward model with a known water content and afterslip parameter and we use this data set as an input data to our inversion scheme. The posterior probability density function not only recovers the target parameters but also informs us of the inherent tradeoffs between the two relaxation mechanisms (Extended Data Fig. 5).

We explore uniform water content C_{OH} in olivine in the range 0.0003 to 0.04 wt% (50 to 6,000 H atoms per million Si atoms), mantle temperatures T_m from 1,350 °C to 1,400 °C, and reference velocities for afterslip in the range $V_0 = 0$ – $2.75 \mu\text{m s}^{-1}$. We explore different values of the transient creep parameters A_K from $A_K = 0$ to $A_K = 3A$, and G_K from $G/2$ to $3G$. We assume all other physical and *in situ* parameters the same for transient creep and steady-state creep. We assume that the laboratory-derived values for the constitutive parameters, which were carried out at much higher strain rate than at typical geological conditions, scale to natural conditions. Geodesy cannot independently constrain the water sensitivity of transient creep and the parameter A_K , so the product $A_K(C_{OH})'$ for transient creep should be considered as a lump parameter. We also investigate a range of strain rates for dislocation creep from 10^{-17} s^{-1} to $2 \times 10^{-14} \text{ s}^{-1}$. Our best-fitting model in Fig. 2 has $V_0 = 1.75 \times 10^{-6} \text{ m s}^{-1}$, $A_K = A/2$, and $G_K = G$. Incorporating transient creep affects the estimate of water content in the mantle, as previously inferred in other studies^{49,50}, by lowering the water content required to fit the data (Fig. 2a). The inferred afterslip parameter V_0 is similar to what is found in other tectonic settings^{25,51}.

We obtain the probability density function of the water content in the asthenosphere by either marginalizing out the afterslip parameter

$$\sigma_{\text{marginal}}(m_1) = \int_0^{\infty} \sigma_{\text{M}}(m) dm_2 \quad (17)$$

where m_1 is the water content in olivine and m_2 is the afterslip parameter; or taking the conditional probability density function around the most likely value of the bivariate distribution

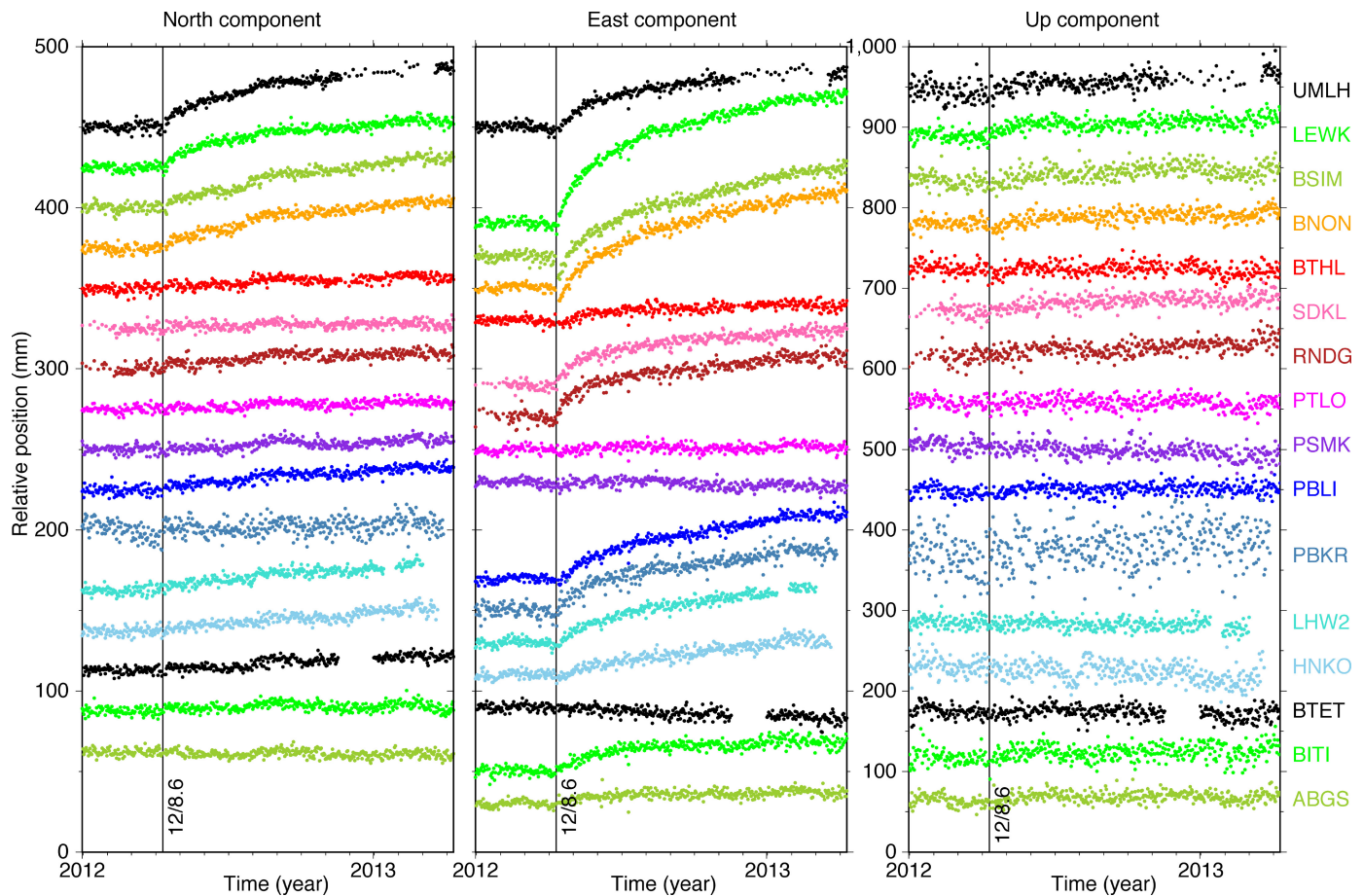
$$\sigma_{\text{conditional}}(m_1) = \frac{\sigma_{\text{M}}(m_1, m_2)}{\int_0^{\infty} \sigma_{\text{M}}(m_1, m_2) dm_1} \quad (18)$$

where m_2 is the most likely value of the afterslip parameter. The posterior probability density function for water content and afterslip in the Wharton basin is shown in Fig. 2a and Extended Data Fig. 6.

The best-fit forward model is shown in Fig. 2 and Extended Data Fig. 7. The small misfit in the GPS time series may be due to our simplifying modelling assumptions, which ignore reactivation of the megathrust or the Sumatran fault and internal deformation of the accretionary prism.

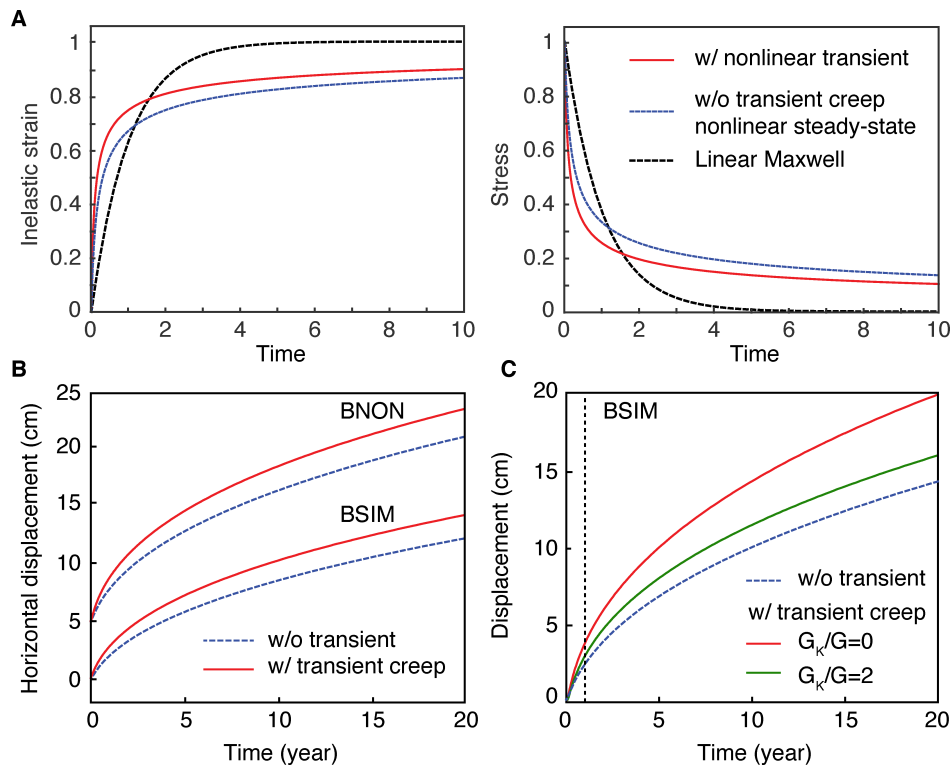
Code availability. The numerical software used in this study is hosted at <https://bitbucket.org> and is available from the corresponding author on request.

31. Feng, L. *et al.* A unified GPS-based earthquake catalog for the Sumatran plate boundary between 2002 and 2013. *J. Geophys. Res.* **120**, 3566–3598 (2015).
32. Ranalli, G. Transient creep in the upper mantle. *Il Nuovo Cimento C* **3**, 405–419 (1980).
33. Hart, E. A phenomenological theory for plastic deformation of polycrystalline metals. *Acta Metall.* **18**, 599–610 (1970).
34. Poirier, J.-P. Shear localization and shear instability in materials in the ductile field. *J. Struct. Geol.* **2**, 135–142 (1980).
35. Chinh, N. Q., Horvath, G., Horita, Z. & Langdon, T. G. A new constitutive relationship for the homogeneous deformation of metals over a wide range of strain. *Acta Mater.* **52**, 3555–3563 (2004).
36. Voce, E. The relationship between stress and strain for homogeneous deformation. *J. Inst. Met.* **74**, 537–562 (1948).
37. Andrade, C. E. N. On the viscous flow in metals, and allied phenomena. *Proc. R. Soc. Lond. A* **84**, 1–12 (1910).
38. Karato, S.-I. *Deformation of Earth Materials: An Introduction to the Rheology of Solid Earth* 34–39 (Cambridge Univ. Press, 2008).
39. Delescluse, M. & Chamot-Rooke, N. Instantaneous deformation and kinematics of the India–Australia plate. *Geophys. J. Int.* **168**, 818–842 (2007).
40. Delescluse, M. *et al.* April 2012 intra-oceanic seismicity off Sumatra boosted by the Banda–Aceh megathrust. *Nature* **490**, 240–244 (2012).
41. Hayes, G. P., Wald, D. J. & Johnson, R. L. Slab1.0: a three-dimensional model of global subduction zone geometries. *J. Geophys. Res.* **117**, B01302 (2012).
42. Barbot, S., Fialko, Y. & Sandwell, D. Three-dimensional models of elasto-static deformation in heterogeneous media, with applications to the eastern California shear zone. *Geophys. J. Int.* **179**, 500–520 (2009).
43. Barbot, S. & Fialko, Y. Fourier-domain Green's function for an elastic semi-infinite solid under gravity, with applications to earthquake and volcano deformation. *Geophys. J. Int.* **182**, 568–582 (2010).
44. Mei, S. & Kohlstedt, D. L. Influence of water on plastic deformation of olivine aggregates: 2. Dislocation creep regime. *J. Geophys. Res.* **105**, 21471–21481 (2000).
45. Jin, D., Karato, S. & Obata, M. Mechanisms of shear localization in the continental lithosphere: inference from the deformation microstructures of peridotites from the Ivrea zone, northwestern Italy. *J. Struct. Geol.* **20**, 195–209 (1998).
46. Behr, W. M. & Platt, J. P. Brittle faults are weak, yet the ductile middle crust is strong: implications for lithospheric mechanics. *J. Geophys. Res.* **41**, 8067–8075 (2014).
47. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation* 1–37 (SIAM, 2004).
48. Sambridge, M. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space. *Geophys. J. Int.* **138**, 479–494 (1999).
49. Freed, A. M., Hirth, G. & Behn, M. D. Using short-term postseismic displacements to infer the ambient deformation conditions of the upper mantle. *J. Geophys. Res.* **117**, 2156–2202 (2010).
50. Pollitz, F. F. Post-earthquake relaxation evidence for laterally variable viscoelastic structure and water content in the Southern California mantle. *J. Geophys. Res.* **120**, 2672–2696 (2015).
51. Wei, S. *et al.* The 2014 M_w 6.1 South Napa earthquake: a unilateral rupture with shallow asperity and rapid afterslip. *Seismol. Res. Lett.* **86**, 344–354 (2015).
52. Chlieh, M. *et al.* Coseismic slip and afterslip of the great M_w 9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97**, S152–S173 (2007).
53. Konca, A. *et al.* Rupture kinematics of the 2005 M_w 8.6 Nias–Simeulue earthquake from the joint inversion of seismic and geodetic data. *Bull. Seismol. Soc. Am.* **97**, S307–S322 (2007).



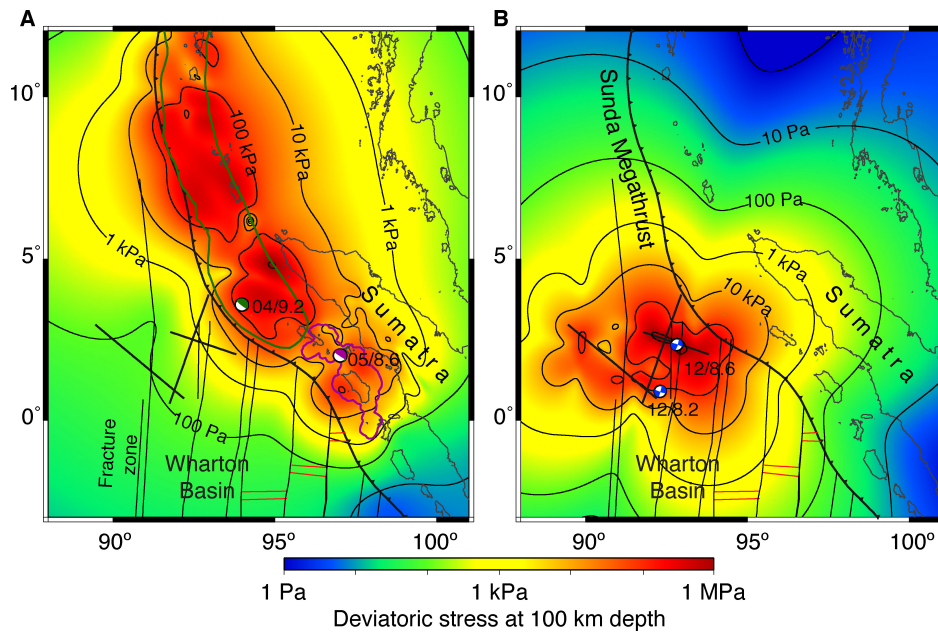
Extended Data Figure 1 | Time series of GPS displacements at the Sumatra GPS Array stations. The effect of previous earthquakes, their postseismic relaxation, and other spurious effects have been modelled and removed from the data (black vertical line labelled '12/8.6' for the year and magnitude of the earthquake). The resulting time series isolate the

postseismic transient following the 2012 M_w 8.6 Indian Ocean earthquake. The (arbitrary) colours indicate the GPS stations. The vertical lines mark the onset of the postseismic transient. The methodology for data preparation is described in refs 6 and 31.



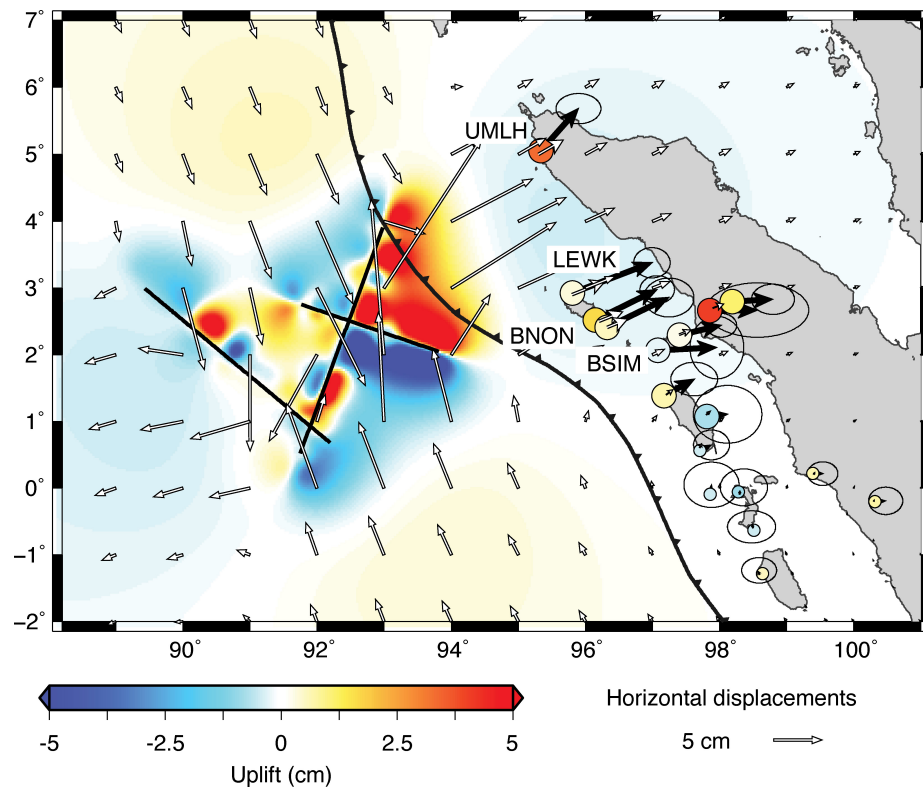
Extended Data Figure 2 | Effect of transient creep of olivine on postseismic transients. **a**, Creep tests. We compare relaxation tests that include the transient creep of olivine (solid red profile) with the response at steady-state for power-law (dashed blue profile) and linear (dashed black profile) viscoelastic materials. The stress perturbation and relaxation times are chosen to highlight typical behaviours. The resulting strain, stress and time are non-dimensionalized. The transient creep accelerates the initial response, but slows down the subsequent relaxation. **b**, Effect of

transient creep on models of postseismic relaxation following the Indian Ocean earthquake. Inclusion of transient creep accelerates the deformation at stations BNON and BSIM, other parameters being the same. **c**, Role of the hardening coefficient G_K on models of postseismic deformation with transient creep at GPS station BSIM. The hardening coefficient controls how much stress is relaxed by transient creep and how long the effect is sustained. The hardening coefficient G_K in our models for the Wharton basin is equal to the background rigidity.

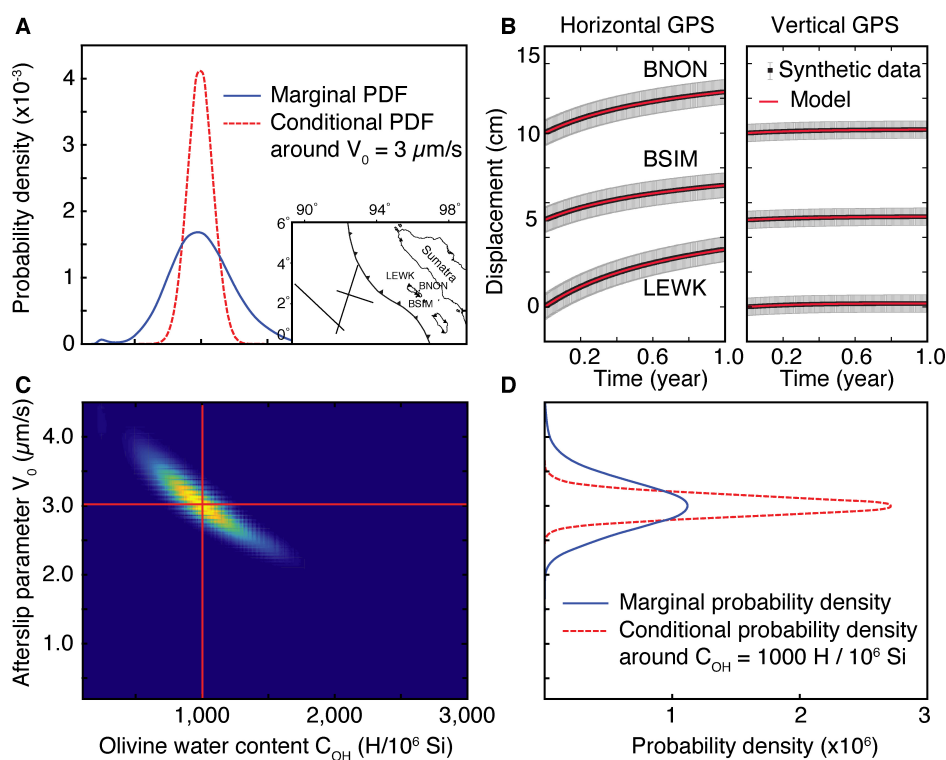


Extended Data Figure 3 | Comparison of the stress changes caused by earthquakes at 100 km depth. Data is from the 2012 M_w 8.6 Indian Ocean⁴ earthquake and the 2004 M_w 9.2 Aceh-Andaman⁵² and the 2005 M_w 8.6 Nias⁵³ earthquakes (coloured beach balls) at 100 km depth. The stress change due to the giant and great Sunda megathrust earthquakes

is one to two orders of magnitude lower than the one due to the 2012 event in the Wharton basin asthenosphere. The negligible effect of the megathrust events is due to their stress changes concentrating below the megathrust and in the mantle wedge. In addition, some of this stress has been relaxed in the many years that separate the events.

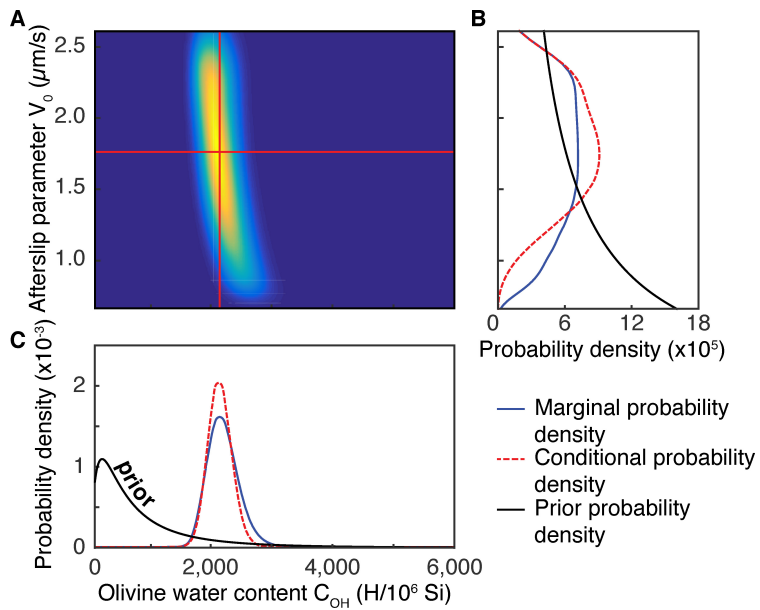


Extended Data Figure 4 | Simulation of the surface postseismic displacements after one year due to stress-driven afterslip on the Indian Ocean coseismic faults. The predicted horizontal displacements (white arrows) are aligned with the GPS observations (black arrows), but the modelled vertical displacements (background colours) are opposite to those measured by GPS (coloured circles).

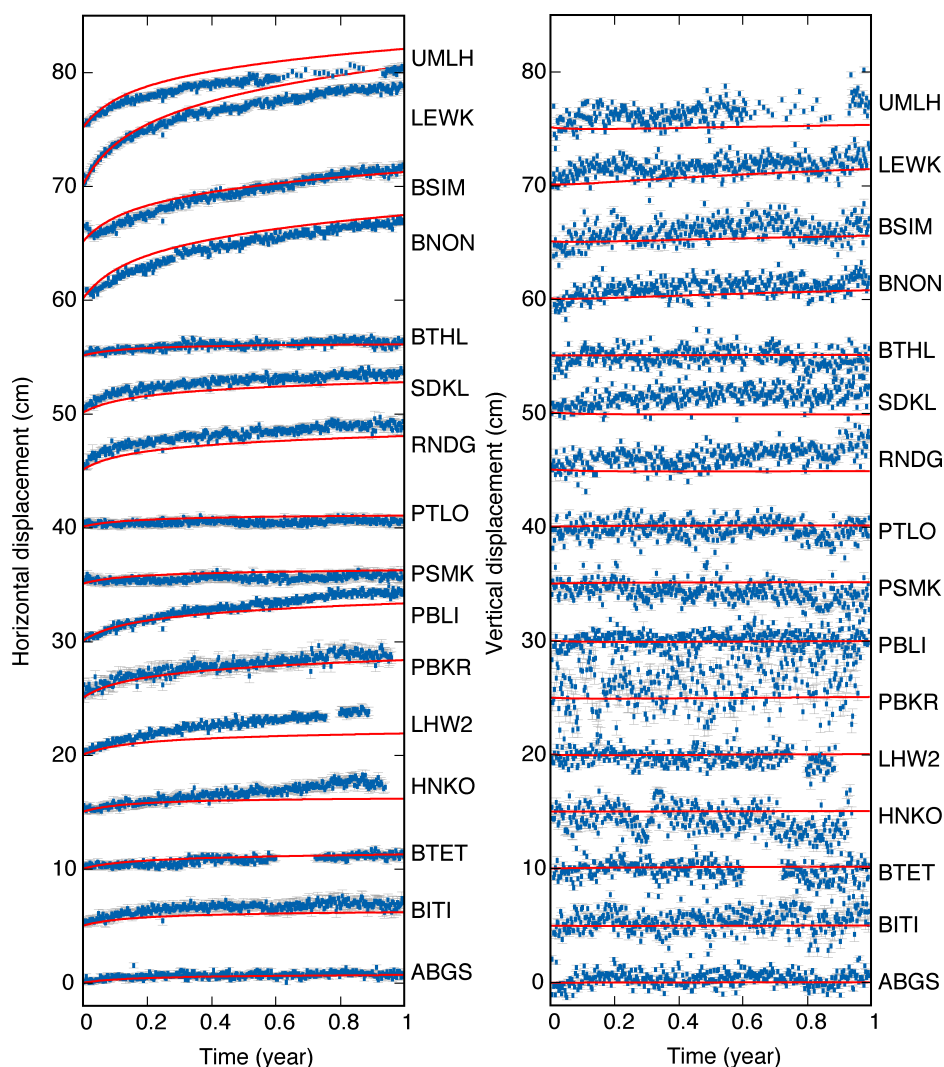


Extended Data Figure 5 | Benchmark of the Bayesian inversion method for water content in olivine. We produce a synthetic data set of GPS time series at the SuGAR network (inset in **a** and black profile with uncertainties in **b**) assuming the afterslip parameter $V_0 = 3 \times 10^{-6} \text{ m s}^{-1}$, a water content of $C_{\text{OH}} = 1,000 \text{ H atoms per million Si atoms}$, and no transient creep. **c**, The bivariate probability density for water content and the afterslip parameter has a maximum at the target value of the model parameters

and documents the tradeoffs between model parameters. **a**, Conditional probability density along the red horizontal profile in **c** and marginal probability density of water content in olivine. **d**, Conditional probability density along the red vertical profile in **c** and marginal probability density of the afterslip parameter. In principle, the spatial distribution of the SuGAR GPS stations with one year of data provides an unbiased estimate of the water content in the asthenosphere.



Extended Data Figure 6 | Probability density of water content in olivine from assimilation of the SuGAR time series of postseismic deformation and prior information from geochemical estimates. **a**, The bivariate probability density of water content and the afterslip parameter. There are few tradeoffs between the two parameters. **b**, Prior (black profile), marginal (blue profile) and conditional (dashed red profile) probability densities of the afterslip parameter. The prior distribution is $1/V_0$, that is, the limit of no prior information for Jeffrey's parameters⁴⁷. **c**, Prior information, marginal and conditional probability densities for water content in olivine. The prior density is a log-normal distribution with a mean value of 600 H atoms per million Si atoms and a standard deviation of 1.



Extended Data Figure 7 | Observed (blue squares with uncertainties) and modelled (red profiles) GPS time series of postseismic displacements after the 2012 M_w 8.6 Indian Ocean earthquake.

The figure shows all the SuGAR time series available and considered in the study. Left, horizontal displacements. Right, vertical displacements. The model reproduces the subtle uplift of the forearc island stations in

the Northern Hemisphere and the temporal evolution of displacements at most stations. The small misfit at some stations (Fig. 2d) can be attributed to the model-simplifying assumptions such as no reactivation of the Sunda megathrust or the Sumatran fault and no internal deformation in the accretionary prism.

A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns

Faiyaz Notta¹, Michelle Chan-Seng-Yue^{1*}, Mathieu Lemire^{1*}, Yilong Li^{2*}, Gavin W. Wilson¹, Ashton A. Connor¹, Robert E. Denroche¹, Sheng-Ben Liang³, Andrew M. K. Brown¹, Jaeseung C. Kim^{1,4}, Tao Wang^{4,5}, Jared T. Simpson^{1,7}, Timothy Beck¹, Ayelet Borgida⁸, Nicholas Buchner¹, Dianne Chadwick³, Sara Hafezi-Bakhtiari^{1,3}, John E. Dick^{1,6,9}, Lawrence Heisler¹, Michael A. Hollingsworth⁸, Emin Ibrahimov¹, Gun Ho Jang¹, Jeremy Johns¹, Lars G. T. Jorgensen¹, Calvin Law¹⁰, Olga Ludkovski⁹, Ilinca Lungu¹, Karen Ng¹, Danielle Pasternack¹, Gloria M. Petersen¹¹, Liran I. Shlush⁹, Lee Timms¹, Ming-Sound Tsao^{4,9}, Julie M. Wilson¹, Christina K. Yung¹, George Zogopoulos¹², John M. S. Bartlett¹, Ludmil B. Alexandrov¹³, Francisco X. Real¹⁴, Sean P. Cleary^{15,16}, Michael H. Roehrl^{3,9}, John D. McPherson^{1,4}, Lincoln D. Stein^{1,6}, Thomas J. Hudson^{1,6}, Peter J. Campbell^{2,17} & Steven Gallinger^{1,15,16}

Pancreatic cancer, a highly aggressive tumour type with uniformly poor prognosis, exemplifies the classically held view of stepwise cancer development¹. The current model of tumorigenesis, based on analyses of precursor lesions, termed pancreatic intraepithelial neoplasm (PanINs) lesions, makes two predictions: first, that pancreatic cancer develops through a particular sequence of genetic alterations^{2–5} (*KRAS*, followed by *CDKN2A*, then *TP53* and *SMAD4*); and second, that the evolutionary trajectory of pancreatic cancer progression is gradual because each alteration is acquired independently. A shortcoming of this model is that clonally expanded precursor lesions do not always belong to the tumour lineage^{2,5–9}, indicating that the evolutionary trajectory of the tumour lineage and precursor lesions can be divergent. This prevailing model of tumorigenesis has contributed to the clinical notion that pancreatic cancer evolves slowly and presents at a late stage¹⁰. However, the propensity for this disease to rapidly metastasize and the inability to improve patient outcomes, despite efforts aimed at early detection¹¹, suggest that pancreatic cancer progression is not gradual. Here, using newly developed informatics tools, we tracked changes in DNA copy number and their associated rearrangements in tumour-enriched genomes and found that pancreatic cancer tumorigenesis is neither gradual nor follows the accepted mutation order. Two-thirds of tumours harbour complex rearrangement patterns associated with mitotic errors, consistent with punctuated equilibrium as the principal evolutionary trajectory¹². In a subset of cases, the consequence of such errors is the simultaneous, rather than sequential, knockout of canonical preneoplastic genetic drivers that are likely to set-off invasive cancer growth. These findings challenge the current progression model of pancreatic cancer and provide insights into the mutational processes that give rise to these aggressive tumours.

Pancreatic cancer will be the second leading cause of cancer-related death in a decade and the biological basis for the aggressive nature of this disease is largely undefined. Motivated by this, we explored the pancreatic cancer genome to address this concern. These genomes are highly unstable¹³, as evidenced by the marked modifications to

the DNA copy number landscape. Although this instability is further exacerbated with metastatic progression¹⁴, it remains unclear when the instability begins relative to the key genetic alterations that give rise to the invasive clone. Also, whether this instability propagates through single copy number changes that accumulate one after another or through large numbers of concurrent changes has not been fully addressed. These questions have important basic and translational implications. As a first step, the mechanisms at the root cause of this instability need to be identified. Mutational phenomena such as chromothripsis and polyploidization have been linked to unstable tumours^{15,16} and aggressive tumour behaviour¹⁷, indicating that they play a role in pancreatic cancer development. These particular phenomena are considered to accelerate cancer evolution because the DNA damage that ensues from such mitotic errors must be resolved in one or few rounds of cell division; otherwise the cell would die. To date, the extensive fibrosis in pancreatic cancer has obstructed the sequencing resolution needed to clearly decipher these events. In this study, we performed an in-depth analysis of more than 100 whole genomes (Extended Data Fig. 1) from purified primary and metastatic pancreatic tumours (referring to ductal adenocarcinoma only), focussing on the mutational phenomena linked to rapid tumour progression.

To evaluate polyploidization, we developed and validated a new informatic tool, termed CELLULOID, which estimates tumour ploidy and copy number from whole-genome data (Fig. 1a and Extended Data Fig. 2). We found that 45% (48/107) of tumours displayed changes in copy number consistent with polyploidization (ploidy solutions can be found in Supplementary Information). Of the polyploid tumours, 88% (42/48) were tetraploid and the rest were hexaploid. The mean ploidy of diploid tumours was 1.95, whereas those tumours that underwent genome duplication and triplication was 3.38 and 5.40 (relative to 4 and 6), indicating that a larger proportion of the genome was lost in the latter subgroup (Extended Data Fig. 3a, b), consistent with previous data¹⁶. Polyploid tumours had higher incidences of mutation in *TP53* ($P = 0.02$, Fisher's exact test; Extended Data Fig. 1e) and harboured 1.5-fold more copy number alterations compared to diploid tumours (median value of 112 versus 77, $P = 0.003$, t -test; Extended Data Fig. 3c).

¹Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. ²Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ³UHN Program in BioSpecimen Sciences, Department of Pathology, University Health Network, Toronto, Ontario M5G 2C4, Canada. ⁴Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada.

⁵Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁶Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁷Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada. ⁸Eppley Institute for Research in Cancer, Nebraska Medical Center, Omaha, Nebraska 68198, USA. ⁹Princess Margaret Cancer Centre, University Health Network (UHN), Toronto, Ontario M5G 2M9, Canada. ¹⁰Division of Surgical Oncology, Sunnybrook Health Sciences Centre, Odette Cancer Centre, Toronto, Ontario M4N 3M5, Canada. ¹¹Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA. ¹²Research Institute of the McGill University Health Centre, Montreal, Québec, Canada, H3H 2L9. ¹³Theoretical Biology and Biophysics (T-6) and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 87545.

¹⁴Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain. ¹⁵Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. ¹⁶Department of Surgery, University Health Network, Toronto, Ontario M5G 2C4, Canada. ¹⁷Department of Haematology, University of Cambridge, Cambridge CB2 0XY, UK.

*These authors contributed equally to this work

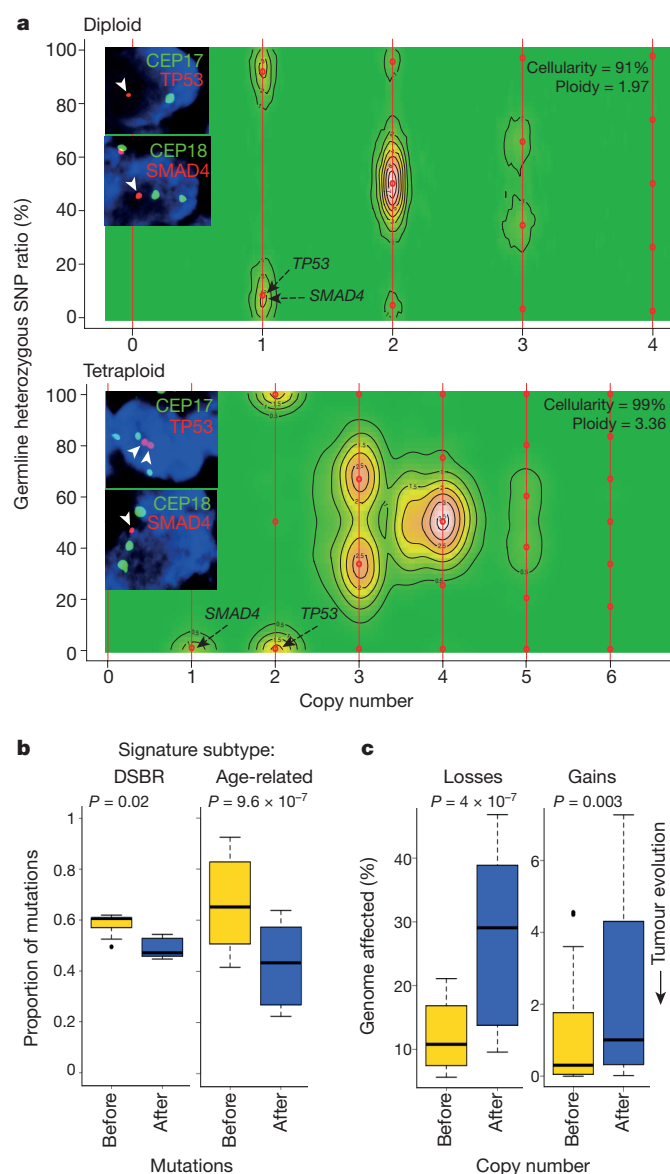


Figure 1 | Polypliodization in pancreatic cancer. **a**, CELLULOID profiles of a diploid (Ashpc_0008) and a tetraploid (Ashpc_0005) tumour. The predicted copy number of *SMAD4* and *TP53* genes is indicated with black arrows. Inset shows a FISH validation of the predicted copy number of *SMAD4* and *TP53* genes. CEP, centromeric probes. **b**, Proportion of mutations that occurred before (yellow) or after (blue) polypliodization. Cases were segregated based on mutational signature subtype: DSBR ($n = 5$; left) and age-related ($n = 32$; right). Owing to the increased genetic instability in polyploid cells, mutations in regions of copy number of 4 in tetraploids were used in this analysis. **c**, Fraction of the genome lost and gained either before (yellow) or after (blue) polypliodization. Box and whisker plots depict median and 10–90 percentile ranges. P values are indicated and were derived using a t -test. A detailed description of these data is given in Supplementary Results.

The marked loss of genomic material relative to baseline ploidy and increased amount of copy number alterations in polyploids demonstrates that these genomes are highly unstable.

We then used mutation data to infer the timing of the polypliodization event in tumour evolution (Supplementary Results). All cases were first categorized according to their dominant mutational signature, since specific aetiologies drive mutation accrual¹⁸. Two subgroups were evident: one where C > T transitions dominated, linked to the process of cytosine deamination (approximately 80% Age-related, Extended Data Fig. 3d) and another where all six classes of base substitutions

were more-or-less balanced—a phenomenon associated with defects in double-strand break repair (DSBR, 17%; Extended Data Fig. 3d). Accordingly, half of the DSBR cases carried germline or somatic mutations in *BRCA1/2* (ref. 13). The remaining cases were comprised of heterogeneous signatures previously identified by Alexandrov *et al.*¹⁸ (Extended Data Fig. 3d).

We found that most mutations preceded polypliodization in both mutational subgroups (Fig. 1b). By contrast, most copy number losses and gains occurred after polypliodization, an effect that was markedly magnified when the size of the copy number change was taken into account (losses: $P = 4.3 \times 10^{-7}$; gains: $P = 0.003$, t -test; Fig. 1c and Extended Data Fig. 3e). This implies that changes in copy number that precede polypliodization were smaller and focal whereas those that come after are larger and more structurally damaging to the genome. Some of these larger changes are likely to be a consequence of the improper segregation of chromosomal material gained during polypliodization. Copy number alterations corresponding to the polypliodization event were commonly seen at integer values and indicate that such events are mostly or fully clonal (CELLULOID solutions in Supplementary Information). Two conclusions emerge from these data: first, polypliodization occurs after an extended diploid phase of mutation accrual; and second, changes in copy number related to polypliodization come to rapidly dominate in the tumour within a shorter timeframe, suggesting they are relevant to disease progression.

Many diploid and polyploid tumours harboured focal copy number alterations that oscillated between a few DNA copy-states, characteristic of chromothripsis¹⁵. We developed a sensitive algorithm, termed ChromAL (see Methods and Supplementary Results), to differentiate chromothripsis from localized gradual events that accumulate over time. We found that 65.4% (70/107) of tumours harboured at least one chromothripsis event (solutions provided in Supplementary Information). A similar frequency was observed in an independent genome cohort (60%, $n = 50$ out of 84, Supplementary Results). Of all chromothripsis events, 11% occurred on chromosome 18 (Extended Data Fig. 4a), resulting in the loss of the key tumour suppressor gene *SMAD4*. By comparing the consensus copy number profiles of tumours with and without chromothripsis, we found that *SMAD4* loss was accompanied by a gain in a region of chromosome 18 that harbours *GATA6*, an oncogene implicated in pancreatic cancer development (Extended Data Fig. 4b, top panel and Supplementary Fig. 1). Furthermore, 8% of events were observed on chromosome 12. The consensus copy number profile of these cases revealed a focal amplification in the region of *KRAS* (Extended Data Fig. 4b, middle panel). These amplifications commonly affected the mutant *KRAS* allele either directly, when chromothripsis and breakage–fusion–bridge (BFB) cycles were combined (Extended Data Fig. 4c, tumour Pcsi_0290), or indirectly, when polypliodization was subsequent to a chromothripsis event that removed the wild-type copy (Extended Data Fig. 4c, Pcsi_0356). There was significantly more chromothripsis in polyploid tumours than in diploid tumours, confirming the greater genetic instability in the former subgroup ($P = 0.013$, Fisher's exact test; Extended Data Fig. 4d). We observed worse overall survival in patients whose tumours had such an event ($P = 0.025$, log-rank test; Supplementary Fig. 2). The high prevalence of chromothripsis in pancreatic cancer, together with previously established links between chromothripsis and aggressive tumour behaviour in other cancers^{15,17}, strongly implicate this mutational processes as a key part of pancreatic cancer development. Notably, these data directly support the ‘catastrophic’ model of pancreatic cancer progression proposed by Real¹⁹ more than a decade ago.

We next performed a series of focused analyses, using individual tumours to illustrate the broad principles of the approach applied to the genome cohort. The data presented above raises an important question: how much of the overall genetic instability in these tumours can be attributed to a single chromothripsis event? In Pcsi_0082, a tetraploid tumour, 63% of all copy number alterations could be attributed to five

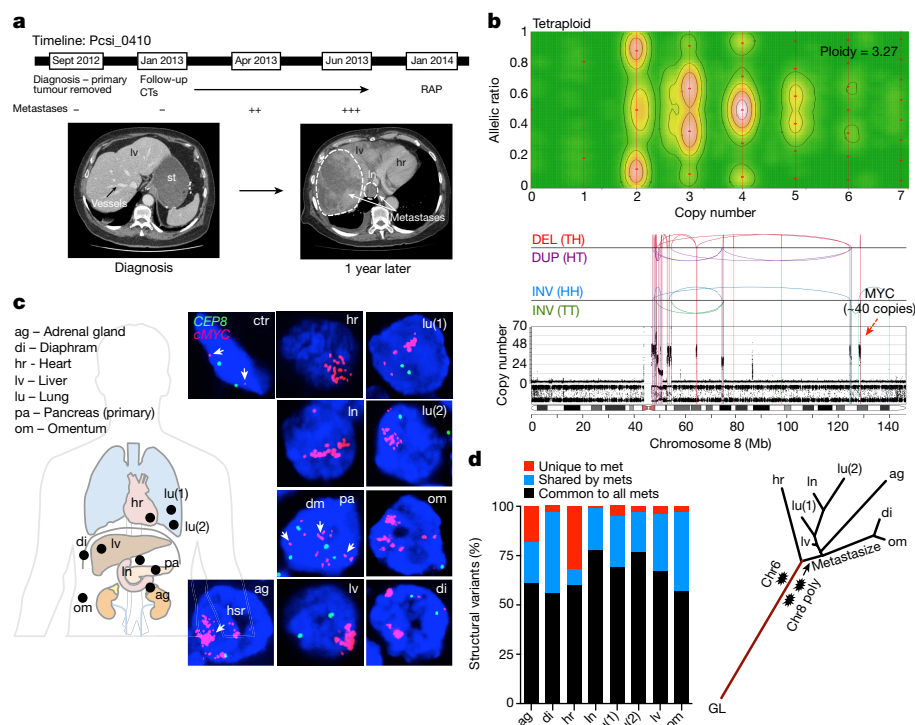


Figure 2 | Chromothripsis and polyploidization in a patient with metastatic progression. **a**, Timeline (top) and computerized tomography scan (CT; bottom) images of Pcsi_0410. White dashed lines indicate metastases. Eight distinct metastases from Pcsi_0410 (see image in **c**) were sequenced. RAP, rapid autopsy. **b**, Polyploidization (top) and chromothripsis (bottom) event from the adrenal gland metastasis (see also Extended Data Fig. 7). **c**, FISH analysis of *MYC* amplification in primary tumour and all metastases. ctr, control (fibroblasts). **d**, Left, the proportion of structural variants common to all (black), shared by two or more

(blue), or unique to each metastasis is shown. Right, copy numbers and structural variants were used to reconstruct phylogenetic tree of metastatic progression. The primary tumour was surgically removed one year before autopsy and fresh-frozen material was not available for whole-genome sequencing. It is possible that branch lengths of the phylogenetic tree would vary if the primary tumour were included in this analysis. Lines are to scale with the copy-number-based clustering dendrogram presented in Supplementary Fig. 15, with the exception of germline origin (GL), which is half the length.

distinct chromothripsis events, on chromosomes 8, 13, 15, 16 and 18 (Extended Data Fig. 5a). As chromothripsis is sustained and resolved in a single cell-division cycle^{20,21}, we can approximate that more than half of the genomic damage in Pcsi_0082 was incurred from approximately five aberrant mitoses. Because Pcsi_0082 had undergone polyploidization, we were able to infer the timing of chromothripsis events relative to the genome doubling using the magnitude of the copy number changes. As chromothripsis occurs on one copy of DNA, the events sustained on chromosomes 13, 16 and 18 must have occurred after polyploidization because the copy number changes on these chromosomes mostly vary by one (Extended Data Fig. 5a, events 2, 4 and 5). By contrast, the chromothripsis on chromosomes 8 and 15 occurred while the tumour was still diploid, since these copy number changes vary in multiples of two, a result of genome doubling (Extended Data Fig. 5a, events 1 and 3). Across all polyploid tumours, we observed that more than half (59%) of all chromothripsis events transpired before polyploidization (ChromAL solutions). This suggests that polyploidization further exacerbates the pre-existing genetic instability in these tumours. Overall, many copy number alterations in pancreatic cancer are acquired through rapid bursts of genetic change from a single or few mitotic events (Extended Data Fig. 5b) rather than a set of gradual events that accumulate over time.

To investigate the role of these mitotic events in disease progression, we analysed the genomes of 15 distinct metastases from six patients (Extended Data Fig. 6 and Supplementary Results). In one case of fulminant metastatic progression (Pcsi_0410), eight distinct metastases were sequenced (Fig. 2a shows the progression timeline). All metastases were polyploid and also carried two distinct chromothripsis events, one on chromosome 6 and another on chromosome 8, that resulted in the marked amplification of *MYC* (20–40 copies), resembling a double

minute (Fig. 2b, c and Extended Data Fig. 7a). The final copy number in areas of loss of heterozygosity (LOH) in both chromothripsis events is two, indicating that both chromothripsis events occurred before polyploidization (Extended Data Fig. 7b). Using fluorescence *in situ* hybridization (FISH), we confirmed that the primary tumour was also polyploid and harboured chromothripsis (Fig. 2c and Supplementary Fig. 3a, b). Thus, we can infer that both chromothripsis events preceded polyploidization and that the systemic spread of the disease occurred after polyploidization by a clone that harboured all three mitotic events (Fig. 2d). An additional chromothripsis event was detected on chromosome 13 in the adrenal gland metastasis (Supplementary Fig. 3c), consistent with previous data on ongoing genetic instability with metastatic progression¹⁴. Overall, we observed that chromothripsis was maintained in metastases if it was present in the primary tumour (Extended Data Fig. 6d). These data support the notion that the majority of genetic instability precedes metastases and is fostered early in tumorigenesis. If the dominant clonal lineage of the primary tumour arises from these types of mitotic events, it suggests that intra-tumoural heterogeneity in pancreatic cancer¹⁰ follows this event, akin to the ‘big-bang’ model proposed for colon cancer²².

The central tenet of the PanIN progression model posits that alterations in *KRAS*, *CDKN2A*, *TP53* and *SMAD4* are acquired as part of a consecutive series of events in tumour evolution. To directly test this model, we used DNA rearrangements to reconstruct the evolutionary history of allelic losses of tumour suppressors based on evidence that allelic alterations are early events in tumorigenesis (Supplementary Results and Luttges *et al.*⁵). Ashpc_0005, a tetraploid tumour, had a complex pattern of rearrangements involving chromosomes 9, 17 and 18, where *CDKN2A*, *TP53* and *SMAD4* are found (Fig. 3a). Several features of this rearrangement pattern facilitate the reconstruction of

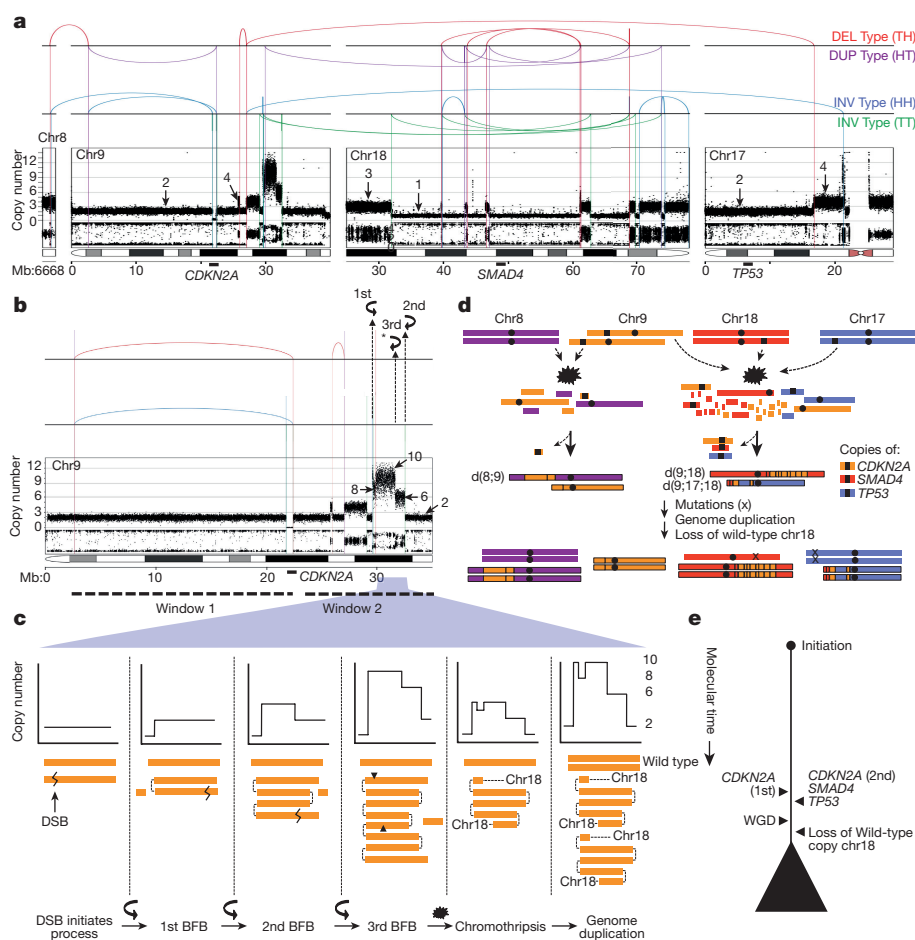


Figure 3 | Simultaneous knockout of pancreatic cancer driver genes.

a, Rearrangement profile of chromothripsis in Ashpc_0005. Positions of key genes (*CDKN2A*, *TP53*, *SMAD4*) are shown at the bottom. DEL, deletion; DUP, duplication; INV, inversion. **b**, Two distinct rearrangement windows on chromosome 9. In window 2, three fold-back inversions (two mapped and one unmapped, marked with an asterisk) are highlighted with curved black arrows. The copy number state of segments as a result of BFB cycles is shown with straight black arrows. **c**, Schematic depiction of the three cycles of BFB that generated the final copy number state in window 2 (**b**).

Zigzag symbol indicates DNA double-strand break to initiate BFB.

d, Temporal order of events based on rearrangement profile. The leftover *TP53* and *SMAD4* alleles carry inactivating mutations (x). As both *TP53* alleles carry the mutations (ploidy > 1), this mutation was acquired before genome duplication. Relative timing of the *SMAD4* mutation cannot be inferred because there is only one copy of this allele and the mutation is fully clonal. d(8;9), d(9;18) and d(9;17;18) refer to candidate derivative chromosomes based on DNA rearrangement profiles. **e**, Summary of tumour evolution in Ashpc_0005. WGD, whole-genome duplication.

the mutational events in this tumour. First, there are two independent sets of rearrangements on chromosome 9 that flank *CDKN2A* (Fig. 3b, windows 1 and 2), indicating that the two copies of this gene were lost as part of independent chromothripsis events. Second, there are distinct amplified DNA segments in window 2 (Fig. 3c) that are bounded by a specific type of rearrangement referred to as a fold-back inversion, an alteration that leaves behind steep copy number drops (>2) indicative of a cycle of BFB¹⁴. Three steep copy number drops in window 2 are evidence of three cycles of BFB (Fig. 3c). Third, the intervening change in copy number (from 10 to 8) on one of these amplified segments suggests that a chromothripsis event followed three cycles of BFB and was likely to be the final major event that stabilized the derivative chromosome²³ (Fig. 3c, penultimate panel). Fourth, all copy number changes in the event are in multiples of two, indicating that polyploidization followed the BFB cycles and chromothripsis (Fig. 3a). Finally, the copy number change on chromosome 18 from 3 to 1 (rather than 4 to 2) indicates that one wild-type copy of this chromosome was lost after polyploidization (Fig. 3a). The relative order of the first and the second copy losses of *CDKN2A* cannot be deciphered, but a single event involving BFB and chromothripsis knocked out a single copy of *CDKN2A*, *TP53* and *SMAD4* in synchronized fashion (Fig. 3d, e). Using rearrangements to reconstruct the

sequence of events in a second case (Pcsi_0171) demonstrated that a single chromothripsis event simultaneously knocked out *CDKN2A* and *SMAD4* (Extended Data Fig. 8). Notably, rearrangement patterns in 16% of cases (17/107) combined allelic alterations in *KRAS*, *CDKN2A*, *TP53* and *SMAD4* genes, predominantly as double knockouts (14% if only tumour-suppressor genes are considered; Supplementary Fig. 4). In a proof-of-principle experiment using single-cell sequencing in a tumour where rearrangements did not span these genes, we found an ancestral clone that harboured a *SMAD4* loss but retained *TP53* and *CDKN2A* (Extended Data Fig. 9). These data provide direct evidence that a number of cases do not conform to the accepted mutational hierarchy predicted by the PanIN progression model and warrant future investigation into the sequence of mutational events that give rise to these aggressive tumours.

Studies dating back two decades have been critical in moulding the current perspective of how pancreatic cancer develops¹. Key features of our data provide a framework to broaden this view. First, analysis of polyploid tumours revealed that most mutations accumulate when these tumours are still diploid. Assuming that preneoplastic cells are diploid, a fraction of these mutations must be preneoplastic. In line with this reasoning, Murphy *et al.* have demonstrated that preneoplasms in pancreatic cancer acquire an extensive mutation burden but remain

non-invasive²⁴. This suggests a prolonged preneoplastic phase predates the onset of invasive disease and that copy number events are crucial for transformation (Extended Data Fig. 10). These data carry implications for the design of future studies on the early detection of pancreatic cancer¹¹. Second, copy number changes from chromothripsis are essentially clonal, suggesting that these events are sustained early in tumorigenesis. The inactivation of well-known preneoplastic drivers (*CDKN2A*, *TP53*, *SMAD4*) *en bloc* strongly supports this notion and implies that chromothripsis can be a transforming event under the right gene context^{17,23}. Our data also raise the possibility that some pancreatic cancers may not progress through a linear series of PanIN lesions¹⁹. Why catastrophic mitotic phenomena are so frequent in pancreatic cancer cannot be easily answered. Perhaps the extensive fibrosis in these tumours, known to suppress tumour development^{25,26}, apply a selective pressure that favours punctuated events over gradual ones. Lastly, pancreatic cancer is well known for its proclivity to metastasize. In mouse models of pancreatic cancer, genetic instability contributes to metastatic progression²⁷. If chromothripsis is indeed the transforming event in some tumours, as our data suggest, a single event could thus confer a cell with both invasive and metastatic properties. In this scenario, there would be a very short latency period between the birth of the invasive clone and the ability of that clone to metastasize^{28,29}. This supposition is consistent with the observation that 80% of pancreatic cancer patients present with advanced disease at diagnosis. How these mutational processes contribute to disease progression and metastatic phenotype is therefore a critical topic of investigation; such knowledge will be essential to guide more effective screening and therapeutic strategies, both for pancreatic cancer and other aggressive tumour types.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2015; accepted 2 September 2016.

Published online 12 October 2016.

- Hruban, R. H., Goggins, M., Parsons, J. & Kern, S. E. Progression model for pancreatic cancer. *Cancer Res.* **6**, 2969–2972 (2000).
- Moskaluk, C. A., Hruban, R. H. & Kern, S. E. *p16* and *K-RAS* gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. *Cancer Res.* **57**, 2140–2143 (1997).
- Wientz, R. E. *et al.* Inactivation of the *p16* (*INK4A*) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. *Cancer Res.* **58**, 4740–4744 (1998).
- Wientz, R. E. *et al.* Loss of expression of *Dpc4* in pancreatic intraepithelial neoplasia: evidence that *DPC4* inactivation occurs late in neoplastic progression. *Cancer Res.* **60**, 2002–2006 (2000).
- Lüttges, J. *et al.* Allelic loss is often the first hit in the biallelic inactivation of the *p53* and *DPC4* genes during pancreatic carcinogenesis. *Am. J. Pathol.* **158**, 1677–1683 (2001).
- Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
- Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–372 (2015).
- Ross-Innes, C. S. *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat. Genet.* **47**, 1038–1046 (2015).
- Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).
- Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
- Chari, S. T. *et al.* Early detection of sporadic pancreatic cancer: summative review. *Pancreas* **44**, 693–712 (2015).
- Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T. J. M.) 82–115 (Freeman, Cooper and Company, 1972).
- Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).

- Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
- Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148**, 59–71 (2012).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Real, F. X. A. A "catastrophic hypothesis" for pancreas cancer progression. *Gastroenterology* **124**, 1958–1964 (2003).
- Zhang, C.-Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
- Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654 (2015).
- Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
- Li, Y. *et al.* Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).
- Murphy, S. J. *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* **145**, 1098–1109.e1 (2013).
- Özdemir, B. C. *et al.* Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* **25**, 719–734 (2014).
- Rhim, A. D. *et al.* Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* **25**, 735–747 (2014).
- Hingorani, S. R. *et al.* *Trp53^{R172H}* and *Kras^{G12D}* cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Haeno, H. *et al.* Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell* **148**, 362–375 (2012).
- Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank N. Simard, S. Zhao and members of the SickKids-UHN Flow facility for technical support. Funding sources for this study include grants to the Pancreatic Cancer Sequencing Initiative program from the Ontario Institute for Cancer Research (OICR), through support from the Ontario Ministry of Research and Innovation, the Canada Foundation for Innovation; research award to F.N. from the OICR and the Canadian Institutes for Health Research (CIHR); Canadian Friends of the Hebrew University, the SMGS Family Foundation, NCI grant P50 CA102701 (Mayo Clinic SPORC in Pancreatic Cancer) and NCI grant R01 CA97075 (Pancreatic Cancer Genetic Epidemiology Consortium). F.N. is supported by a fellowship award from CIHR and is a recipient of a scholar's research award from the Ontario Institute of Cancer Research (OICR), through support from the Ontario Ministry of Research and Innovation. G.Z. is a Clinician–Scientist of the Fonds de la Recherche en Santé du Québec. P.J.C. is a Wellcome Trust Senior Clinical Fellow. T.J.H., L.D.S., J.D.M. and S.G. are recipients of Senior or Clinician–Scientist Awards from the Ontario Institute for Cancer Research.

Author Contributions Data analysis and interpretation was performed by F.N., M.L., Y.L., M.C.-S.-Y., G.W.W., A.A.C., F.X.R., P.J.C., S.G. and T.J.H.; tumour enrichment by S.-B.L., I.L. and F.N.; pathological assessment by T.W., M.-S.T., J.M.S.B., M.H.R. and S.H.-B.; genomics by R.E.D., A.M.K.B., K.N., J.C.K., L.T., N.B., D.P., L.H., E.I., G.H.J., J.J., L.G.T.J., J.D.M., L.D.S., L.I.S., L.H., J.E.D., C.K.Y., T.B. and L.B.A.; FISH by O.L.; CELLULOID analysis by M.L. and single-cell analysis by G.W.W., J.T.S. and F.N. Sample acquisition, annotation and collection from institutes external to University Health Network. was performed by G.M.P., M.A.H., G.Z. and C.L. Sample acquisition, annotation and collection from the University Health Network was performed by J.M.W., A.B., S.G. and S.P.C. The study was designed by F.N., T.J.H. and S.G.; F.N. prepared and wrote the manuscript; the manuscript was edited by M.L., F.X.R., J.E.D., P.J.C., T.J.H. and S.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F. N. (faizyaz.notta@oicr.on.ca) or T. J. H. (tom.hudson@oicr.on.ca).

Reviewer Information *Nature* thanks S. Chanock, M. Rossi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Ethical approval and sample acquisition. A total of 107 surgically resectable samples of pancreatic ductal adenocarcinoma tissue were obtained from collaborating hospitals in Canada and the United States from patients that gave informed consent under the ICGC protocol. 84 samples were obtained from the University Health Network (Toronto, Canada), 14 samples from the Mayo Clinic, 3 samples from the University of Nebraska as part of a rapid autopsy program, 5 samples from Sunnybrook Health Sciences Centre (Toronto, Canada), and 1 sample from McGill University (Montreal, Canada). Consent for WGS was obtained locally at each institute. At the Ontario Institute for Cancer Research, approval was obtained through the University Health Network Research Ethics Board (08-0767-T) and University of Toronto Research Ethics Board (30024). Pre-operatively, blood samples were collected for germline DNA. Where blood was not collected, duodenal mucosa or other non-cancerous tissue was collected post-operatively to obtain germline DNA. Tumours were removed to confirm the diagnosis of ductal adenocarcinoma and pieces were snap-frozen in liquid nitrogen and stored at -80°C or -150°C before proceeding with laser capture microdissection (LCM). For 21 cases (17 UHN, 4 Sunnybrook), fresh tumour material was dissociated and viably sorted at -150°C (below). We obtained clinical follow-ups on the majority of cases.

Sample dissociation and cell sorting. Freshly resected tumours were minced into fine pieces in 10-cm tissue culture dishes using a razor blade. After mechanical dissociation, 9 ml of RPMI supplemented with 1% FBS was added. 1 ml of $10\times$ collagenase/hyaluronidase mix (Stemcell technologies) was added to bring the volume to 10 ml and the sample was placed in a 37°C incubator. Every 20 min, the tissue pieces in the culture dish were pipetted through narrowing orifices (for example, a 10 ml then 5 ml then 1 ml pipette) for a total of 60–120 min. The sample was then passed through a 70–150- μm nylon mesh, centrifuged and resuspended in DMSO (Sigma) based cryopreservation media (20% FBS/10% DMSO final) and placed at -150°C for long-term storage.

For cell sorting, frozen vials of viable cells were thawed via dropwise addition of RPMI solution (IMDM + 20% FBS + DNaseI). Final concentration of DNaseI (Roche Applied Science, 10104159001) in RPMI solution was $200\mu\text{g ml}^{-1}$. After thawing, cells were spun at a low r.p.m. ($\sim 1,000$) for 20 min at 4°C . After spinning, the thawing solution was removed and cells were resuspended in $100\mu\text{l}$ of PBS + 5% FBS for antibody staining and cell sorting. The following antibodies were used for cell sorting: GlyA FITC (BD bioscience, clone HIR2), CD140b PE (BD bioscience, clone 28D4), CD45 PC5 (Beckman Coulter, clone IM1833), EpCAM PerCP-eFluor710 (eBioscience, clone 1B7), CD31 PC7 (eBioscience, clone WM-59), CD90 (BD Biosciences, clone 5E10), CD34 APC7 (BD bioscience, clone 581, custom conjugation). Cell sorting was performed on the BD FACSaria III using 4-laser configuration.

Laser capture microdissection. Snap-frozen tumour tissue embedded in optimal cutting temperature compound was cut into $8\mu\text{m}$ sections and mounted on PEN-Membrane Slides (Leica). Sections were stained with diluted haematoxylin to distinguish tumour epithelium from stroma. A staff pathologist marked tumour sections and LCM was performed according to manufacturer's protocol on the Leica LMD7000 system. Specimens were collected by gravity, contact-free and contamination-free, and directly placed in DNA lysis buffer.

Whole-genome sequencing was performed on DNA from tumour-enriched material. Details of sequencing protocols are included in the Supplementary Methods.

CELLULOID: evaluation of tumour cellularity, tumour ploidy, and absolute copy number profiles. After alignment, reads are counted in 1-kb bins using functions from the R package 'HMMcopy'. These counts are then adjusted for the GC content of each bin using LOESS (local) regression and scaled to the mean (scaled GC-corrected read count (SRC)). Segmentation of the data in both tumour and normal tissue (say, from matched non-malignant tissue or from blood) is performed using penalized least squares, as implemented in the R package 'copynumber'. Each segment is assigned the mean SRC value, which is calculated from the bins within the segment. SRC is proportional to the mean number of chromosomes (copies), averaged over all sequenced cells.

Germline heterozygous positions are extracted in the autosome, except in regions of the genome where duplication or deletion events are observed in the normal tissues. The number of reads supporting each allele (the reference allele—the one observed on the reference human assembly—and the alternate allele) is recorded from the tumour data and the allelic ratio (AR; the proportion of reads supporting the reference allele) calculated. Each heterozygous position is also paired with the SRC value of the segment it belongs to, evaluated from the tumour

data, to form pairs of values (SRC, AR). These pairs of points are represented in a three-dimensional graph as a contour (elevation) plot (Fig. 1). This figure is a visual representation of the autosomal-wide copy-number profile of the tumour. Each peak (or pair of peaks since the graph is reflected around $\text{AR}=0.5$) corresponds to a specific copy number state that summarizes both the total copy number (on the x axis, once appropriately scaled) and the ratio of relative abundance of maternal and paternal copies (on the y axis, once contamination from normal tissues—or tumour cellularity—is accounted for). The relative positions of these peaks can be mathematically derived in the following way.

Let us define the autosomal ploidy of a sequenced sample (that includes both tumour and possibly contamination from normal cells) as:

$$P = \frac{1}{N_B} \sum_b c_b$$

where c_b represents the mean number of chromosomal copies at base b , averaged over all cells, and N_B is the number of autosomal bases. This can be interpreted as the relative abundance of autosomal DNA in the sequenced sample compared to a normal (reference) haploid autosomal genome. We aim to use the SRC values to estimate the ploidy. Re-writing the above as:

$$P \approx K \times \frac{1}{N_{\text{bins}}} \sum_{\text{bin}} \text{SRC}_{\text{bin}}$$

(where K is a scaling constant) is not informative since the SRC values are scaled and relative, making this expression trivial. However, because SRC are scaled to the mean, bins that fall in regions of exactly P copies (averaged over all cells) are expected to display SRC values of 1. Let S be the value of SRC that would be expected in regions where all cells display 2 copies of chromosomes (such regions do not need to actually exist in the sequenced sample). Because of proportionality, we have the relationship:

$$P = \frac{2}{S}$$

thus, ploidy can be evaluated by finding S .

Consider the more general case of a sequenced sample that consists of a proportion n of normal cells and t of tumour cells ($n+t=1$). Because ploidy may differ in normal and tumour cells, these percentages are not equivalent to percentages of reads originating from normal or tumour cells. Consider a segment in the genome that is present in 2 copies in the normal cells and an average of T copies in the tumour cells. The tumour cells can be further broken down in subclones, in proportions t_1, t_2, \dots ($t=t_1+t_2+\dots$), each subclone displaying a different number of copies (T_1, T_2, \dots). Then, by proportionality, the SRC of bins in that segment are expected to take the value:

$$\frac{S}{2} \times (2n + T_1t_1 + T_2t_2 + \dots) \quad (1)$$

To determine the expected AR of heterozygous positions in that segment, the number of copies need to be further broken down into number of maternal and paternal copies: $T_i = M_i + P_i$. Normal cells are assumed to have one maternal chromosome and one paternal chromosome. In a segment that displays M_i maternal and P_i paternal copies in subclone i , the AR is expected to take the value:

$$\frac{n + M_1t_1 + M_2t_2 + \dots}{2n + T_1t_1 + T_2t_2 + \dots} \quad (2)$$

if, say, the maternal chromosome carries the reference allele, and reflected around 0.5 otherwise. Let:

$$\text{EP}(S, n, t, M_1, P_1, M_2, P_2, \dots)$$

represent the (x, y) coordinates described in equations (1) and (2). Let $\text{OP} = \{\text{OP}_i\}$ be the set of observed contour plot peaks (or subset of peaks deemed of particular interest by the user). The algorithm used to estimate S , n and t finds parameters that minimize the total distance between the observed peaks and the expected peak (EP) coordinate closest to each. In other words, if:

$$d_i(S, n, t) = \min_{M_1, P_1, M_2, P_2, \dots} |\text{OP}_i - \text{EP}(S, n, t, M_1, P_1, M_2, P_2, \dots)|,$$

then the algorithm consists on finding S , n and t that minimize:

$$\sum_i d_i(S, n, t).$$

In practice, the number of expected peak locations grows exponentially with the number of subclones and the number of maternal/paternal configurations. The algorithm further depends on a set of allowed copy number configurations (a set of M_i and P_i) that can be set by the user. For example, the user might want to ignore configurations where the number of maternal chromosomes is smaller than the number of paternal chromosomes in one subclone but higher in another; this would reduce the number of possible ARs. Other restrictions may include situations where the number of copies between different subclones cannot differ (by difference or by ratio) by more than some specified threshold.

The objective function to be minimized is not convex and multiple local minima exist. Optimization is done either by simulated annealing if a global minimum is desired (using the R package GenSA) or using the R built-in function 'optim' with grid-defined starting points to survey and inspect a set of local minima.

Once values for S , n and t are obtained, the ploidy in the tumour cells (P_T) can then be calculated as:

$$P_T = \frac{P - 2n}{1 - n}$$

where $P = 2/S$ is the ploidy of the whole sample that was sequenced. The SRC values can be rescaled into their corresponding integer copy number in tumours using equation 1 above.

The above describes the current implementation of an R package named CELLULOID, which can be obtained from <http://github.com/mathieu-lemire>.

Chrom-AL: detecting catastrophic mitotic events. Chrom-AL is an in-house tool developed to standardize the detection of complex rearrangement patterns linked to chromothripsis²⁰. Chrom-AL applies a series of statistical tests and thresholds at the level of the chromosome and also within the windows of the structural events to infer a call. We inspected 80 genomes manually and estimate that the false-positive and false-negative rate of Chrom-AL is ~7% and ~8%, respectively, in our dataset. The tool is designed based on the chromothripsis criteria presented by Korbel and Campbell³⁰. Complex rearrangement patterns can often involve multiple distinct types of mitotic errors (for example, FoSTeS, MMBIR) including a chromothripsis event^{20,21,23}. Chrom-AL is not designed to distinguish chromothripsis from other replication-based mitotic errors, which can also be catastrophic within one or few cell divisions. As such, we use the term chromothripsis to broadly refer to a 'one-off' mitotic catastrophe.

As chromothripsis events typically increase the number of structural variants in a genome, there is a correlation between tumours with increased numbers of structural variants and rate of chromothripsis. Thus, proper structural variant calling becomes a critical parameter in implementation of any algorithm to call chromothripsis. Despite this correlation, a high rate of structural variants does not necessarily imply a chromothripsis event. Thus, the false-positive and false-negative rates of Chrom-AL will probably vary with the overall rate of structural variants that differs amongst tumour types. For this reason, visual inspection still remains a critical tool in evaluating such events. Chrom-AL does not detect chromothripsis events that are predominately driven by a single type of structural variation. For example, on rare occasions we observed the typical copy number oscillation hallmark of chromothripsis that was connected mostly by head-to-head (HH) or tail-to-tail (TT) inversions. Whether such rearrangements were indeed accumulated over time or all at once is not known. To remain consistent with the criteria discussed below, we excluded these events from the analysis. Below, we describe the criteria and conditions used to detect cataclysmic events by Chrom-AL. Chrom-AL was implemented in R.

Threshold for number of structural variants and copy number alterations at the (chromosomal level); test 1. Catastrophic events typically have large numbers of structural variations and copy numbers changes. Only events with at least 7–8 structural variants and 8 copy number segments were considered in the analysis.

Clustering of break points (chromosomal level); test 2. Catastrophic events are typically localized to particular genomic regions that can be assessed statistically. To do this, we ordered the break points sequentially and calculated the distances between each break point. The distribution of distances was compared against the exponential distribution as described by Korbel and Campbell³⁰ using a Kolmogorov–Smirnov (KS) test and followed by Bonferroni correction. Regions with a $q \leq 0.1$ were considered to display evidence of break-point clustering.

Chromosomal break-point enrichment (chromosomal level); test 3. We observed several instances where structural variants comprising a catastrophic event were scattered chromosome-wide and did not cluster within a particular region of a chromosome. Thus, they failed the KS test described above. To account for this shortcoming, we performed an additional test to determine if structural variants were enriched on any particular chromosome than would be expected by chance. To identify chromosomes enriched for structural variants, a hypergeometric test was run on each chromosome based on all the breakpoints identified

in the tumour. This was followed by a Bonferroni correction. Chromosomes with a $q \leq 0.1$ were identified as having a high rate of break points.

Join distribution (chromosomal and window level); test 4. In paired-end sequencing, all structural variants can be categorized into four read-pair orientations based on the direction of the + or – reads: tail-to-head (+/–, TH), head-to-head (–/–), tail-to-tail (+/+) or head-to-tail (–/+, HT). Pairs in standard orientation (+/–) are considered to be a deletion-type structural variant with a TH join. Duplication-type structural variants are in the reverse-orientation –/+ and defined by a HT join. Inversions can be both in the forward (+/+) orientation or reverse (–/–) orientation. In the forward orientation, they were defined as TT and in the reverse orientation they were defined as HH. Using read-pair information for structural variants, we classified each structural variant based on their segment joins. In a catastrophic event, we expect structural variants of all four types to be present. For each region we tested this hypothesis. To initially run the test, we required at least one type of read-pair join from each of the four subtypes to be present. A multinomial test, from the EMT v1.1 package, was run to test the distribution of segment joins against an equal distribution. The regions with $P > 0.05$ were considered to show evidence of equal distributions of segment joins.

Copy-number oscillations (chromosomal and window level); test 5. Catastrophic events typically display oscillations in copy number that vary between a few states. However, when chromothripsis is co-opted with BFB cycles as part of a single catastrophic event, there will be some segments in the event that will oscillate between limited copy number states and other segments that may appear to increase in a stepwise manner. To be categorized as a bona fide one-off event, there must be some sequential segments that retain an oscillation pattern. We required at least 4 sequential segments in any catastrophic event must oscillate between two different states. Due to polyploidization, the amplitude of the copy number step was defined as variable (1, 2 or more).

Interspersed LOH (chromosomal and window level); test 6. Chromothripsis drives copy number losses, and thus copy number oscillations should correspond to interspersed loss of heterozygosity (LOH). To test for LOH, we identified all the high confidence germline heterozygous SNPs in the genome and determined the allelic ratio in the tumour sample. The distributions of allelic ratios between each sequential copy number segments were compared using a t -test. A minimum of ten positions had to be identified within each copy number segment to be processed otherwise those segments were excluded from the analysis. A Bonferroni correction test was run. Those segments in which $q \leq 0.1$ were considered significantly different. To show evidence of interspersed LOH, at least four comparisons had to be made (thus at least five copy number segments had to be present in the region). At least 50% of the compared segments had to show some significant difference in the distribution of allele ratio to be classified as showing interspersed LOH.

Chromosome-level analysis. Genomic regions were first evaluated at the chromosome level. For each sample, all chromosomes were independently evaluated for the above tests. For tests 2 and 3, we used copy number break points for segments where a matching structural variant could not be mapped. The importance of this point is shown in Extended Data Fig. 5c (bottom left panel; Ashpc_0008, event 2). In this case, there was a chained chromothripsis event connecting chromosomes 3 and 20. On chromosome 3, the left edge at 42.8 Mb was part of the chromothripsis event but the corresponding structural variant to this copy number loss is not mapped. This was also the case for the right edge of the chromothripsis event on chromosome 20 (7.1 Mb). In this scenario, utilization of the copy number break point was critical in the tests to decipher whether this was indeed a chromothripsis event. If copy number break points are not integrated into the analysis, such events would go undetected or be misclassified. We found that including the copy number break point was necessary to properly establish the DNA windows of chromothripsis events, especially when structural variants could not be properly mapped (discussed below).

Identification of DNA rearrangement windows. The next step was to identify the borders of the catastrophic event on each chromosome. Catastrophic events typically display overlapping structural variants throughout the region of the event. To localize the chromosomal window where the catastrophic event occurred, we selected the left and right borders of overlapping structural variant break points. Structural variants resulting in translocations were used to establish the rearrangement window when at least two independent translocations were detected between the same two chromosomes. In this manner, we could establish inter- and intra-chromosomal windows to facilitate the segregation of multi-chromosome events from single-chromosome events. Each window was flanked with 6 kb on either end. The windows that define each candidate catastrophic event were used for downstream analysis.

Window-level analysis. A window was first scored on whether there were at least eight structural variants present within the window. Each window was then evaluated for tests 4–6.

Classification of single-chromosome versus multi-chromosome catastrophic events. Single-chromosome catastrophic events were classified when all structural variants within a window occurred on the same chromosome. In the case of translocations, at least two structural variants had to have occurred between the same two chromosomes to be considered a multi-chromosome event.

Event (criterion 1 versus criterion 2). Each window was independently scored. If a window was classified as a catastrophic and was involved in a multi-chromosome event, both windows on either side of the translocation were considered to be catastrophic at the chromosomal level but were counted as a single catastrophic event. Through a large number of iterations, in which the tests described above were iteratively optimized, we established two distinct criteria: 'maximize sensitivity' and 'specificity of detection'.

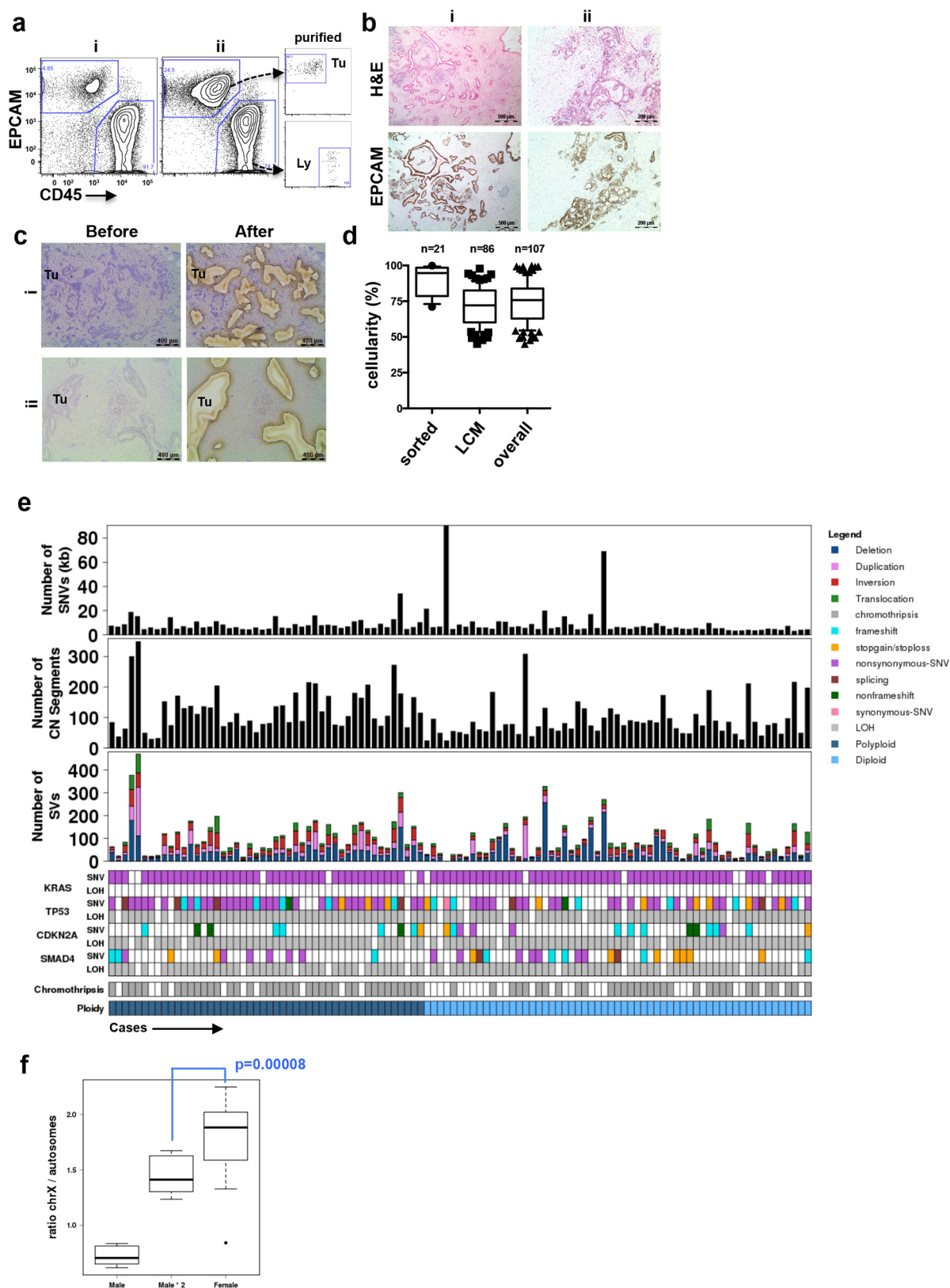
Criterion 1 (CR1). To be classified as an event under CR1, a region had to pass at least five of the six chromosomal level tests (test 1–6). A window had to be

identified with at least eight structural variants and the window had to pass the segment-join and the interspersed LOH test (test 4 and 6).

Criterion 2 (CR2). To be classified as event under CR2, a region had to pass at least 5 of the following tests: the 6 chromosomal level tests (test 1–6), the identification of a window with at least 8 structural variants, the window segment join and the window interspersed LOH test (test 4 and 6). In addition to these conditions, the window had to have at least 7 structural variant events and had to pass window oscillation (criterion 5).

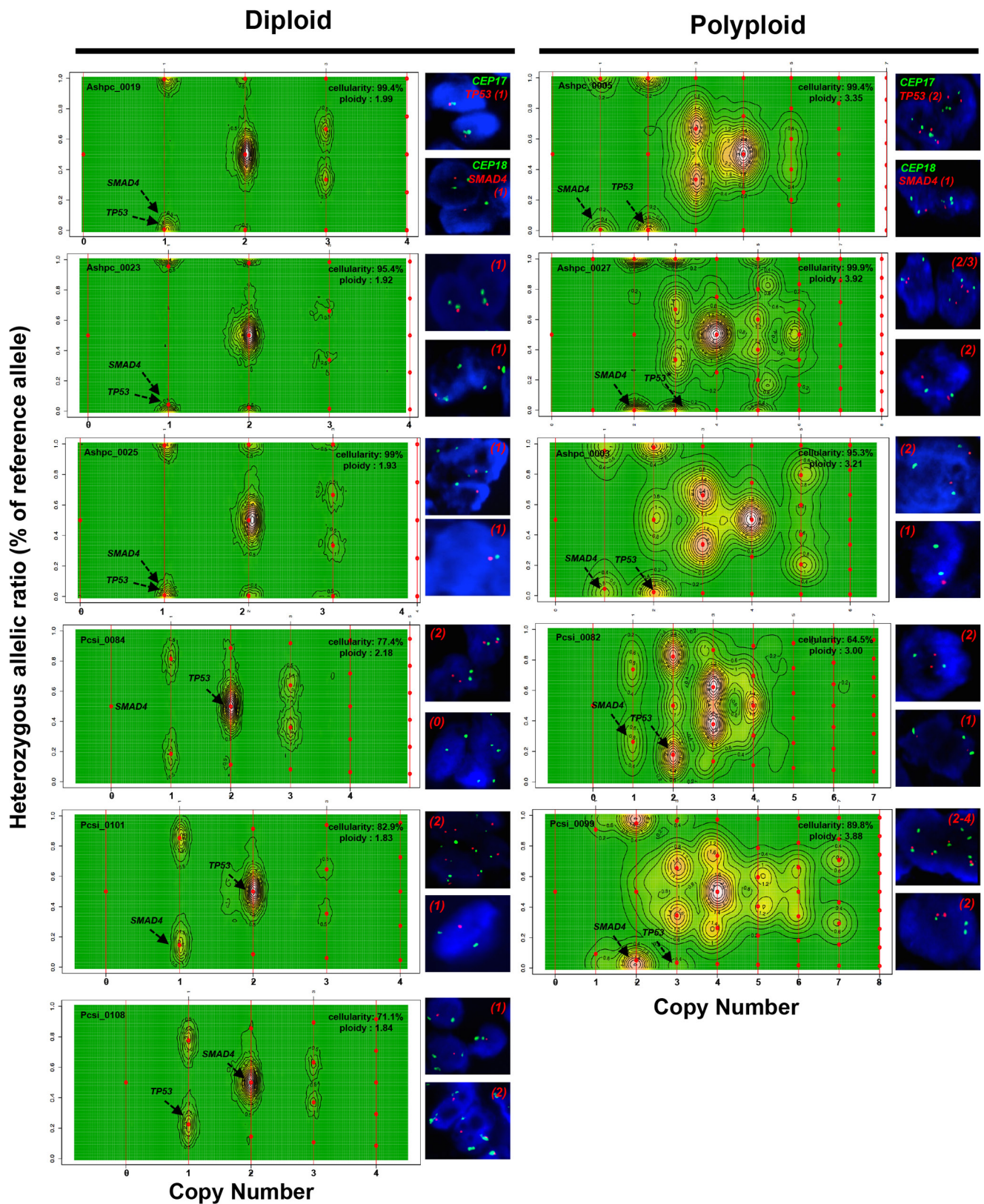
Data availability. Raw data (fastq files) and clinical information on the patient cohort are available from the International Cancer Genome Consortium (ICGC) data portal at <http://dcc.icgc.org>. DNA sequencing data have also been deposited in the European Genome-phenome Archive (EGA): EGAD00001001956.

30. Korbelt, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).



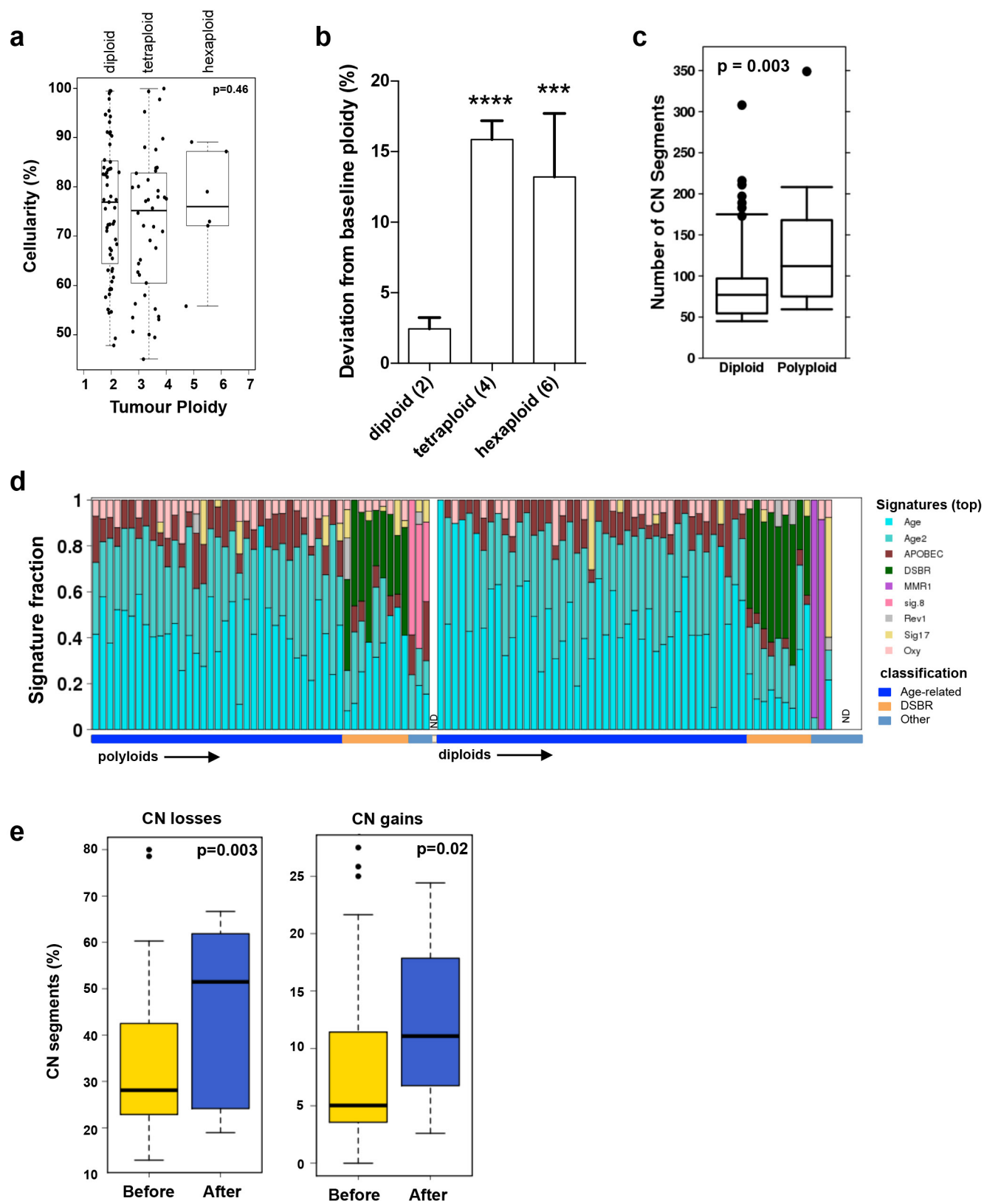
Extended Data Figure 1 | Tumour enrichment and overview of somatic alterations in the cohort used in this study. **a**, Flow cytometry profiles of EpCAM and CD45 from 2 (of 21) representative cases of pancreatic ductal adenocarcinoma (PDA) (i, ii). On the right, post-sort analysis of EpCAM⁺ cells (Tu) and CD45⁺ lymphocytes (Ly) demonstrates the high level of purity obtained from flow-sorting. **b**, Immunohistochemical analysis of formalin-fixed tumours using the EpCAM clone for flow sorting in a (H&E, haematoxylin and eosin). Two independent cases are shown (i, ii). **c**, Profiles of haematoxylin-stained sections of PDA before and after LCM from two representative cases (of 86) (i, ii). **d**, Box and whisker

plots represent median and 10–90 percentile ranges of tumour cellularity of flow-sorted ($n = 21$), LCM ($n = 86$) and the total cohort ($n = 107$) of tumours. Dashed line depicts cellularity of bulk tumours that have not undergone enrichment. **e**, Overview of somatic alterations of the cohort used in the study. **f**, X-chromosome mutation ratio in diploid pancreatic cancer genomes showing hypermutation on this chromosome in females. Males were corrected for single copy of the X chromosome by doubling the raw value. P values were derived from t -tests. A more detailed description of these data is provided in Supplementary Results.



Extended Data Figure 2 | CELLULOID validation. The copy number for common alterations (*TP53*, *SMAD4*; shown by black arrow) was derived from ploidy estimates generated by CELLULOID. Six diploid and five polyploid tumours were analysed by FISH (shown on the right of each contour plot). In all cases, the copy number from CELLULOID ploidy estimates was confirmed. In Pcsi_0084 (diploid), CELLULOID predicted zero copies of *SMAD4*. The allelic ratio in this region was 50%

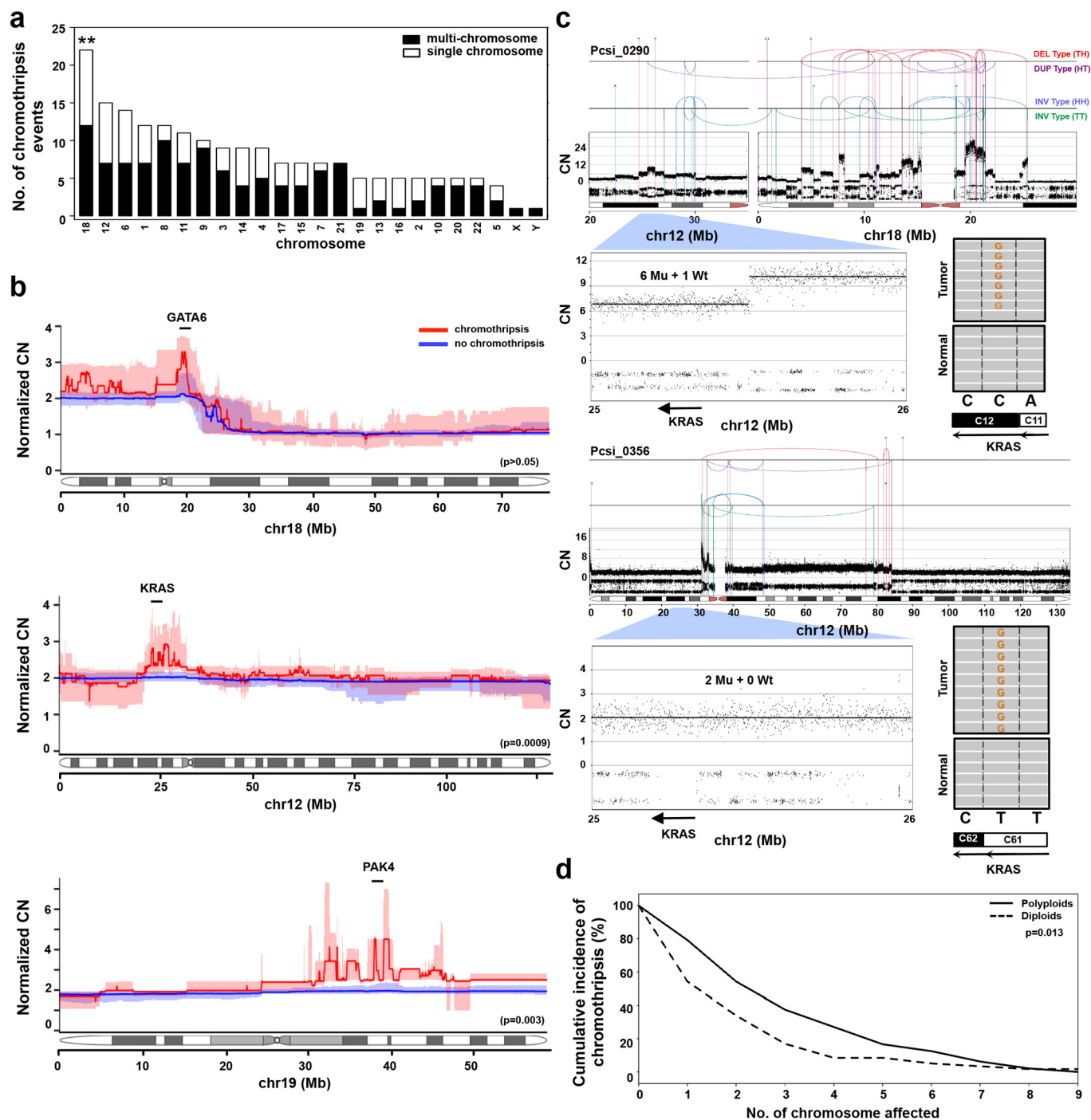
(heterozygous) as only reads from normal cells spanned this region. In Ashpc_0027, both CELLULOID and FISH indicate that this tumour is polyploid. The CELLULOID plot demonstrates that there is a further subclonal amplification in *TP53* from polyploid clone (copy state = 3.2 derived from one allele). FISH analysis shows tumour cells with two or three copies of *TP53* supporting this is subclonal. Copy number by FISH for *SMAD4* and *TP53* is indicated in red at the top right of each plot.



Extended Data Figure 3 | See next page for caption.

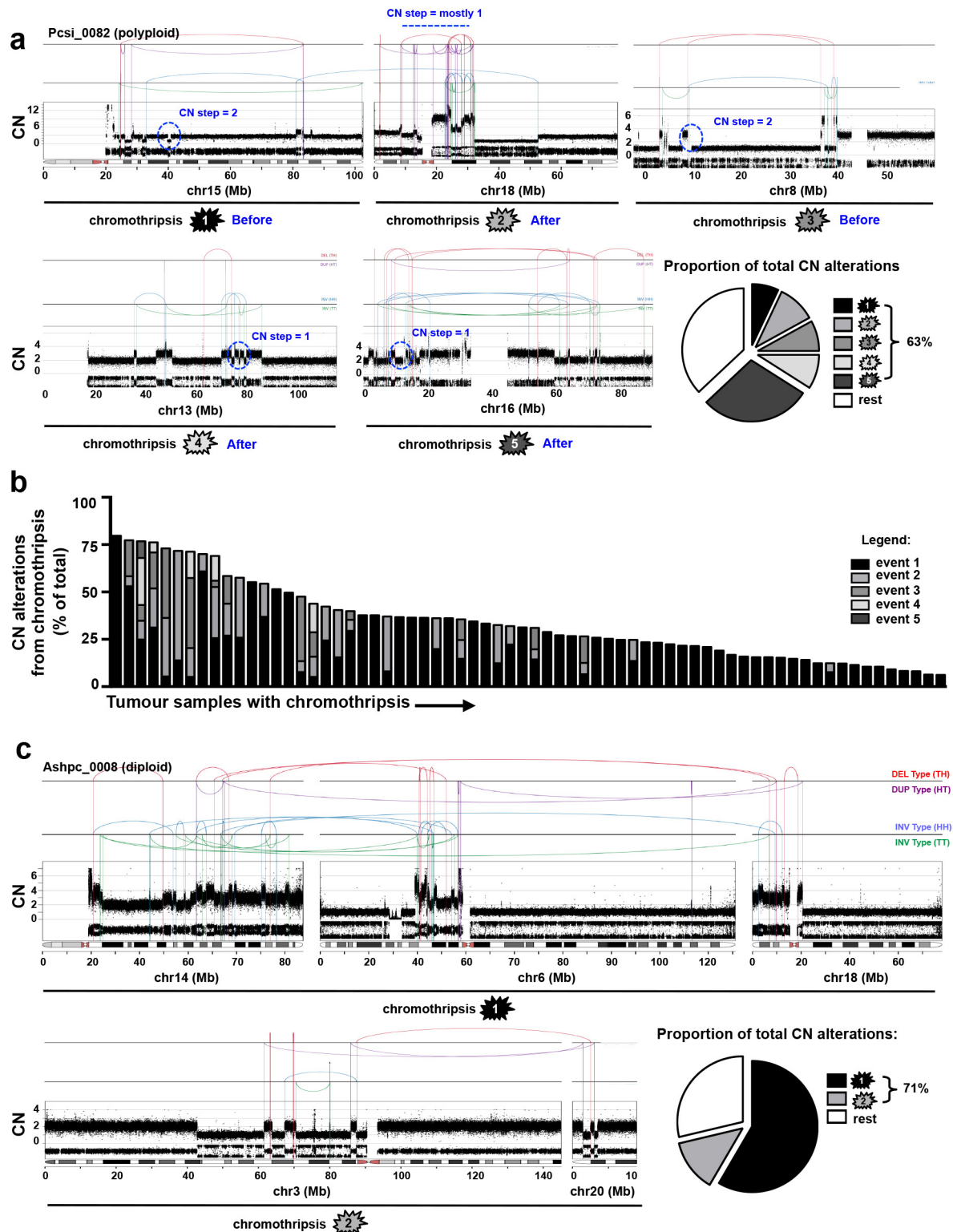
Extended Data Figure 3 | Tumour ploidy and genetic instability in pancreatic cancer. **a**, Tumour ploidy and sample cellularity estimates are interconnected: although the ploidy of a tumour can always be doubled and still provide copy number segments at integer levels (albeit only at even values), the estimate of cellularity would have to decrease. To maintain an allelic ratio at a given value, the proportion of tumour cells has to be reduced to compensate for the higher copy numbers in them (from a cellularity value t to a value $t/(2 - t)$ in the case of a doubling of the ploidy). A test can thus be designed to verify that ploidy estimates have not been systematically over- or underestimated, simply by comparing the distribution of cellularity estimates stratified by ploidy. P value was derived using Kruskal–Wallis test. **b**, Deviation from baseline ploidy in diploids, tetraploids and hexaploids indicates a marked loss of genomic material in polyploids. **c**, Box and whisker plots (showing the median

and 10th–90th percentile ranges) of the total copy number alterations in polyploid and diploid tumours. **d**, Mutational signatures of the 107 genomes used in this study. The signatures were derived using the trinucleotide mutation context as previously published¹⁸. The proportion of individual signature operative in each tumour is shown in the bar plot. The overall classification of each case is indicated below. Signatures of polyploidy tumours is shown on the left, diploids is shown on the right. ND, not done; $n = 1$ polyploid and 4 diploid patient samples. Detailed analysis of mutational signatures in PDA is covered elsewhere (Connor *et al.*, manuscript under review) **e**, Percentage of copy number losses (left) and gains (right) that occurred before (yellow) or after (blue) genome duplication for each polyploid tumour. Box and whisker plots depict median \pm 10th–90th percentile range. P values were derived using a t -test.



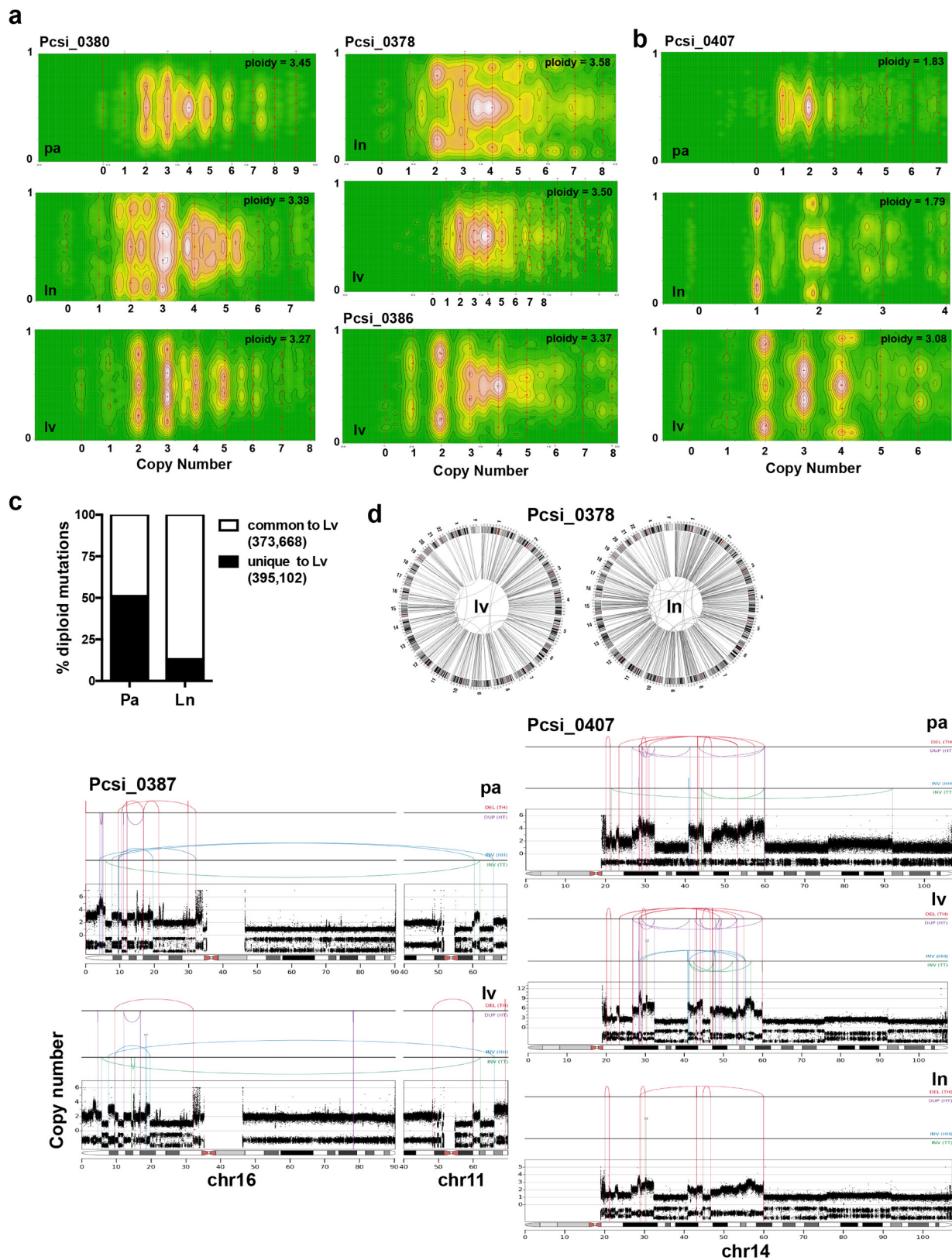
Extended Data Figure 4 | Characterization of chromothripsis events in pancreatic cancer. **a**, The distribution of chromothripsis events across the genome (single-chromosome, white; multi-chromosome, black). $^{**}P < 0.001$ (Monte Carlo sampling, Supplementary Methods). **b**, The specific effects of chromothripsis on the copy number of chromosome 18 (top, $n = 22$), chromosome 12 (middle, $n = 15$), and chromosome 19 (bottom, $n = 5$). Statistical differences in copy number between the groups were performed using Wilcoxon test using 10-kb bins that covered *GATA6* (chromosome 18), *KRAS* (chromosome 12) and *PAK4* (chromosome 19) genes (description of *PAK4* event is covered in supplementary results). Copy number profiles of polyploids were adjusted according to tumour

ploidy to allow comparison against diploids (referred to as 'Normalized copy number' on the y axis). Interquartile ranges for chromothripsis cases are indicated in pale red and for non-chromothripsis cases in pale blue. **c**, Two cases of chromothripsis resulting in the amplification of the mutant *KRAS* allele. In Pcsi_0290, the mutant allele was amplified as part of a multi-chromosomal event involving chromothripsis and BFB with chromosome 18 (top). In Pcsi_0356, the chromothripsis event was co-opted with cycles of BFB to knock out the wild-type allele (bottom). The absolute copy number of the locus encompassing *KRAS* and mutation is shown for each case. **d**, Cumulative incidence of chromothripsis events in polyploid and diploid tumours ($P = 0.013$, Fisher's exact test).



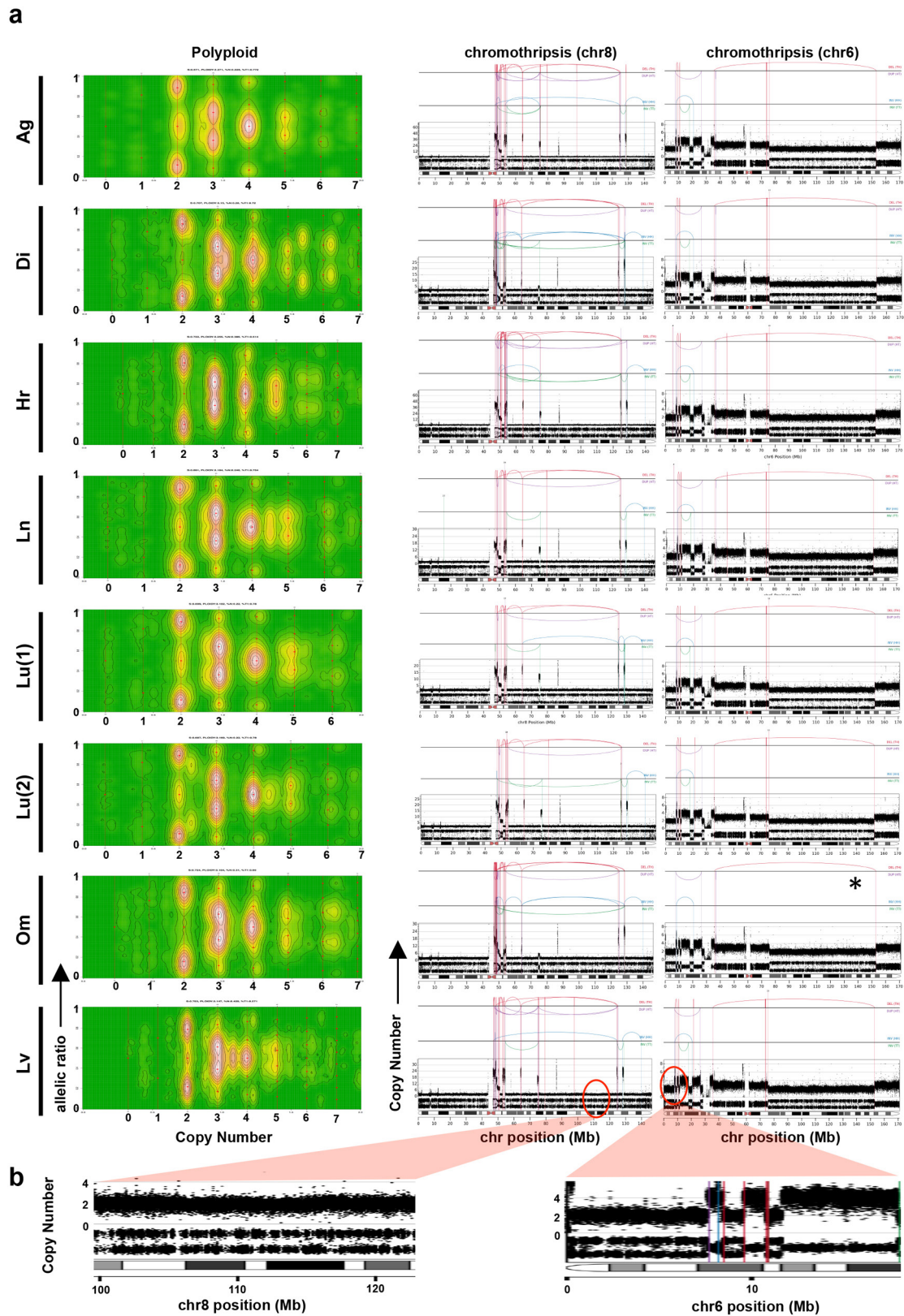
Extended Data Figure 5 | Most copy number alterations arise from individual chromothripsis events. **a**, In Pcsi_0082, five distinct chromothripsis events on chromosome 15 (top, 1), chromosome 18 (top, 2), chromosome 8 (top, 3), chromosome 13 (bottom, 4), and chromosome 16 (bottom, 5) are displayed. Copy number steps on chromosome 15 (1), chromosome 8 (2) are 2 or greater indicating that these events occurred before polyploidization. Single copy number steps on chromosome 18 (2), chromosome 13 (4) and chromosome 16 (5) indicate that these events were sustained after polyploidization. The single rearrangement between

chromosome 15 and chromosome 18 appears to be independent from the chromothripsis on chromosome 18. Pie charts depict the proportion of copy number alterations derived from each chromothripsis event. **b**, Distribution of copy number alterations due to chromothripsis for all cases where such an event was detected by ChromAL. **c**, In Ashpc_0008, two multi-chromosomal chromothripsis events, joining chromosome 14, chromosome 6, chromosome 18 (top, 1), and chromosome 3, chromosome 20 (bottom, 2), are shown (discussed in Supplementary Results).

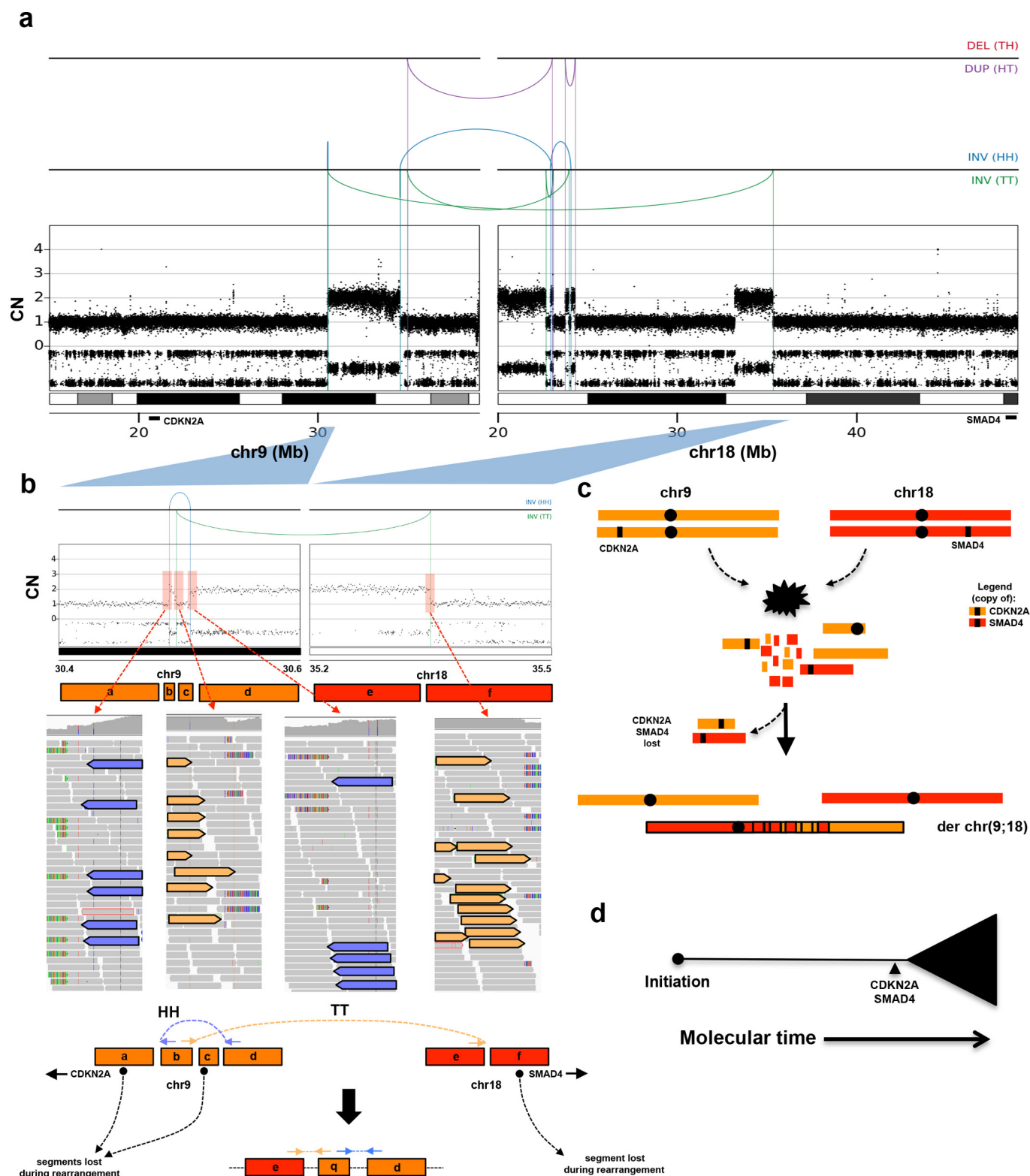


Extended Data Figure 6 | Characterization of chromothripsis and polyploidy in metastases. a, CELLULOID plots illustrating polyploidy in metastases. In Pcsi_0380, the primary tumour was directly available for analysis. Similarly to Pcsi_0378, multiple metastases were polyploid suggesting the primary tumour was also polyploid. The primary tumour was unavailable for sequencing in this case. **b,** A case (Pcsi_0407) with

discordant ploidy amongst different metastases. **c,** Percentage of diploid mutations from liver metastases that are shared (white) or unique (black) when compared to the primary tumour or the lymph node metastasis. **d,** Plots of chromothripsis events in metastases. ln, lymph node; lv, liver; pa, primary tumour.

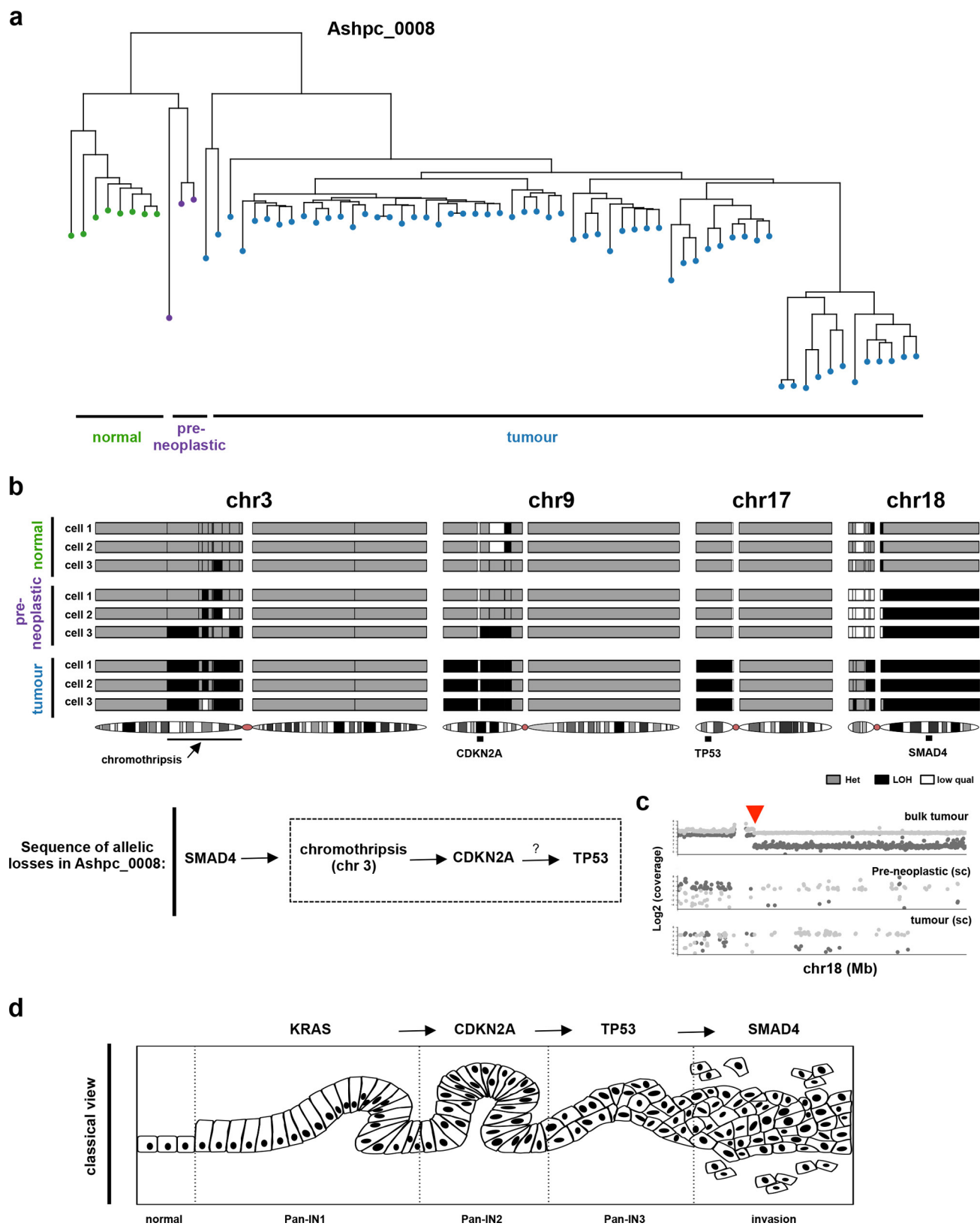


Extended Data Figure 7 | Chromothripsis and polyploidy in Pcsi_0410. This figure accompanies Fig. 2. **a**, CELLULOID (left panel) and chromothripsis plots (middle and right) of the different metastases from a patients with fulminant metastatic progression. **b**, Copy number and LOH from chromosome 8 (left) and chromosome 6 (right) chromothripsis events indicate that these events were sustained before polyploidization.



Extended Data Figure 8 | Case of a simultaneous loss of *CDKN2A* and *SMAD4* due to a chromothripsis event. a, Rearrangement and copy number profile of a multi-chromosome chromothripsis event between chromosome 9 and chromosome 18 (Pcsi_0171). b, Detailed view of the two inversions (one in the head-to-head orientation (HH), the other in tail-to-tail orientation (TT) for more detail, see Methods) in the

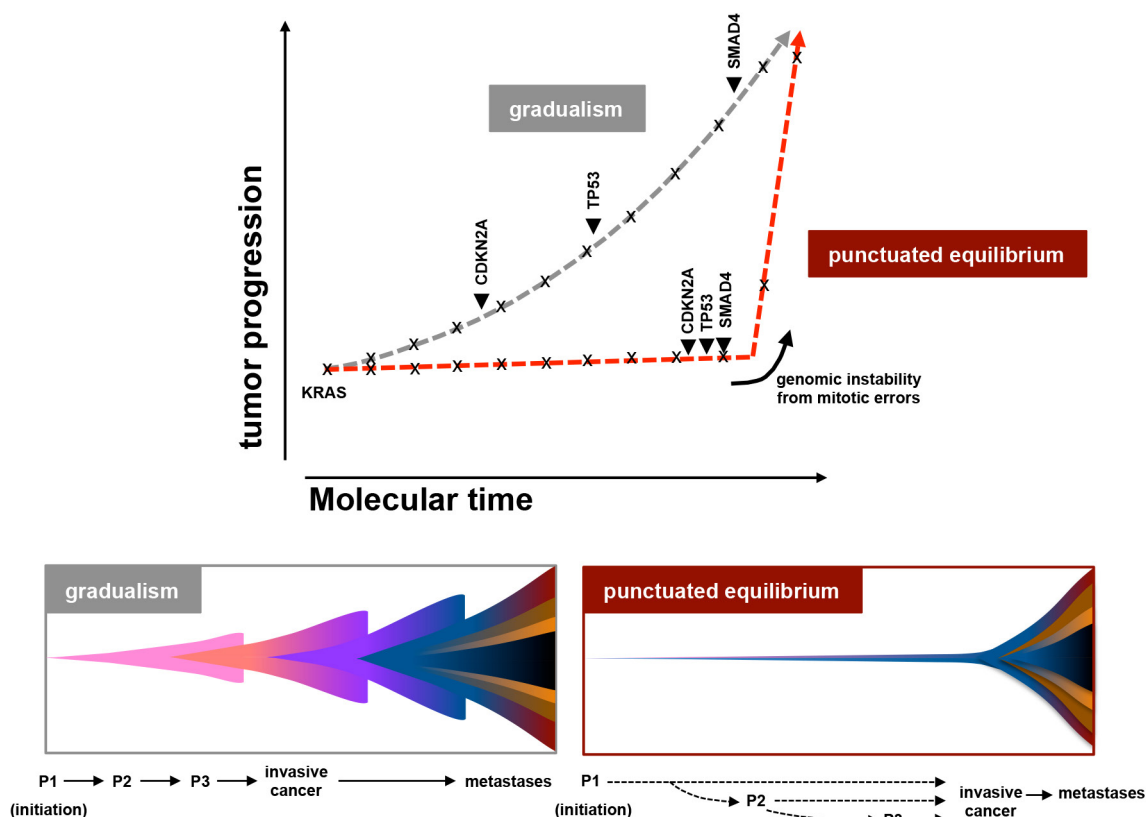
chromothripsis event that resulted in the concurrent loss of *CDKN2A* and *SMAD4*. c, Schematic depiction of the temporal order of events derived from the rearrangement profile shown in a. d, Summary of tumour evolution in Pcsi_0171. A more detailed description of Pcsi_0171 is provided in Supplementary Results.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Single-cell sequencing reconstruction of the evolutionary events when rearrangements did not span the classical pancreatic cancer drivers. **a**, A fresh tumour specimen (Ashpc_0008) was dissociated and single tumour cells were deposited using flow sorting. The whole genomes of 96 single cells were amplified using REPLI-g and paired-end whole-genome sequencing was performed using an Illumina HiSeq 2500 system. Single cells were sequenced to a median whole-genome depth of $3.9\times$ (Supplementary Fig. 18). Only cells with enough whole-genome coverage ($n = 70$) were used in the analysis. This sequencing depth allowed us to track heterozygous SNPs across the whole genome in single cells. Using this methodology, we were able to follow LOH events across the whole genome in single cells that show high concordance with

bulk tumour tissue (Supplementary Fig. 18). Hierarchical clustering based on LOH events across the whole genome was performed and found four independent cell clusters. **b**, Specific LOH events on chromosome 3, chromosome 9, chromosome 17 and chromosome 18 are shown from single cells in **a**. The chromothripsis event on chromosome 3 is shown in greater detail in Extended Data Fig. 5c. A summary of the sequence of allelic losses is shown below. Supportive data that allelic losses precede mutational inactivation is shown in Supplementary Figs 13, 14. **c**, Plot of the shared chromosomal break point on chromosome 18 on the bulk (top), preneoplastic single cell (middle) and tumour single cell (bottom). **d**, The classical model of pancreatic tumour progression.



Extended Data Figure 10 | Theoretical model of pancreatic cancer tumour progression. Shown is the classical model of tumour evolution driven at a gradual pace (grey) and an alternate model driven at punctuated equilibrium (red). In the classical model, there is a period of latency between the driver mutations that lead to tumour development and the multiple, independent, transforming events are required for tumour development (top, grey dashed line; bottom-left schematic). In the punctuated equilibrium model, tumour development can be divided into two major events, the cancer-initiating event and cancer-transforming

event (top, red dashed line; bottom-right schematic). Under this model, most mutations (indicated by \times) would accrue in an extended phase of preneoplastic tumour development. Transformation, probably due to genetic instability from copy number changes (arrow heads) ensuing from a cataclysmic event, would rapidly lead to invasive cancer and metastases. Classical drivers (*KRAS*, *CDKN2A*, *TP53*, *SMAD4*) from the PanIN progression model are overlaid onto these models. Theoretical PanIN stages are shown as P1–P3.

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature20164

Erratum: A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns

Faiyaz Notta, Michelle Chan-Seng-Yue, Mathieu Lemire, Yilong Li, Gavin W. Wilson, Ashton A. Connor, Robert E. Denroche, Sheng-Ben Liang, Andrew M. K. Brown, Jaeseung C. Kim, Tao Wang, Jared T. Simpson, Timothy Beck, Ayelet Borgida, Nicholas Buchner, Dianne Chadwick, Sara Hafezi-Bakhtiari, John E. Dick, Lawrence Heisler, Michael A. Hollingsworth, Emin Ibrahimov, Gun Ho Jang, Jeremy Johns, Lars G. T. Jorgensen, Calvin Law, Olga Ludkovski, Ilinca Lungu, Karen Ng, Danielle Pasternack, Gloria M. Petersen, Liran I. Shlush, Lee Timms, Ming-Sound Tsao, Julie M. Wilson, Christina K. Yung, George Zogopoulos, John M. S. Bartlett, Ludmil B. Alexandrov, Francisco X. Real, Sean P. Cleary, Michael H. Roehrl, John D. McPherson, Lincoln D. Stein, Thomas J. Hudson, Peter J. Campbell & Steven Gallinger

Nature **538**, 378–382 (2016); doi:10.1038/nature19823

In this Letter, owing to a typesetter error the ‘received date’ was incorrectly shown as ‘8 August 2015’ instead of ‘27 November 2015’ in both the print and PDF versions; this has been corrected online.

Cortico–fugal output from visual cortex promotes plasticity of innate motor behaviour

Bao-hua Liu^{1,2}, Andrew D. Huberman³ & Massimo Scanziani^{1,2,4}

The mammalian visual cortex massively innervates the brainstem, a phylogenetically older structure, via cortico–fugal axonal projections¹. Many cortico–fugal projections target brainstem nuclei that mediate innate motor behaviours, but the function of these projections remains poorly understood^{1–4}. A prime example of such behaviours is the optokinetic reflex (OKR), an innate eye movement mediated by the brainstem accessory optic system^{3,5,6}, that stabilizes images on the retina as the animal moves through the environment and is thus crucial for vision⁵. The OKR is plastic, allowing the amplitude of this reflex to be adaptively adjusted relative to other oculomotor reflexes and thereby ensuring image stability throughout life^{7–11}. Although the plasticity of the OKR is thought to involve subcortical structures such as the cerebellum and vestibular nuclei^{10–13}, cortical lesions have suggested that the visual cortex might also be involved^{19,14,15}. Here we show that projections from the mouse visual cortex to the accessory optic system promote the adaptive plasticity of the OKR. OKR potentiation, a compensatory plastic increase in the amplitude of the OKR in response to vestibular impairment^{11,16–18}, is diminished by silencing visual cortex. Furthermore, targeted ablation of a sparse population of cortico–fugal neurons that specifically project to the accessory optic system severely impairs OKR potentiation. Finally, OKR potentiation results from an enhanced drive exerted by the visual cortex onto the accessory optic system. Thus, cortico–fugal projections to the brainstem enable the visual cortex, an area that has been principally studied for its sensory processing function¹⁹, to plastically adapt the execution of innate motor behaviours.

Although the OKR is innate, it is also plastic^{7–11}. Indeed, impairment of the vestibulo–ocular reflex, an innate oculomotor behaviour that works with the OKR to stabilize retinal images, leads to a compensatory increase in the amplitude of the OKR^{11,16–18}. The compensatory increase in OKR relative to the vestibulo–ocular reflex is a striking example of how reflexes are plastically adjusted relative to each other to ensure appropriate motor behaviour. The adaptive plasticity of the OKR is classically studied by impairing the vestibulo–ocular reflex using lesions of the vestibular organ^{11,14,16,17}. We used this experimental paradigm to determine whether the visual cortex is involved in the ensuing compensatory increase in OKR amplitude and to identify the underlying neural circuits.

We elicited a horizontal OKR (here referred to as OKR) in head-fixed adult mice by displaying on a virtual drum a vertical grating that drifted along the azimuth in an oscillatory manner; at the same time, we monitored the right eye with a camera (Fig. 1a, b and Extended Data Fig. 1a–d; Methods). We computed the gain of the OKR as the amplitude of the eye trajectory normalized to the amplitude of the grating trajectory (Methods; Extended Data Fig. 1). The OKR gain depended on the spatial and oscillation frequencies of the grating and varied from animal to animal^{17,20} (Fig. 1c and Extended Data Fig. 1g).

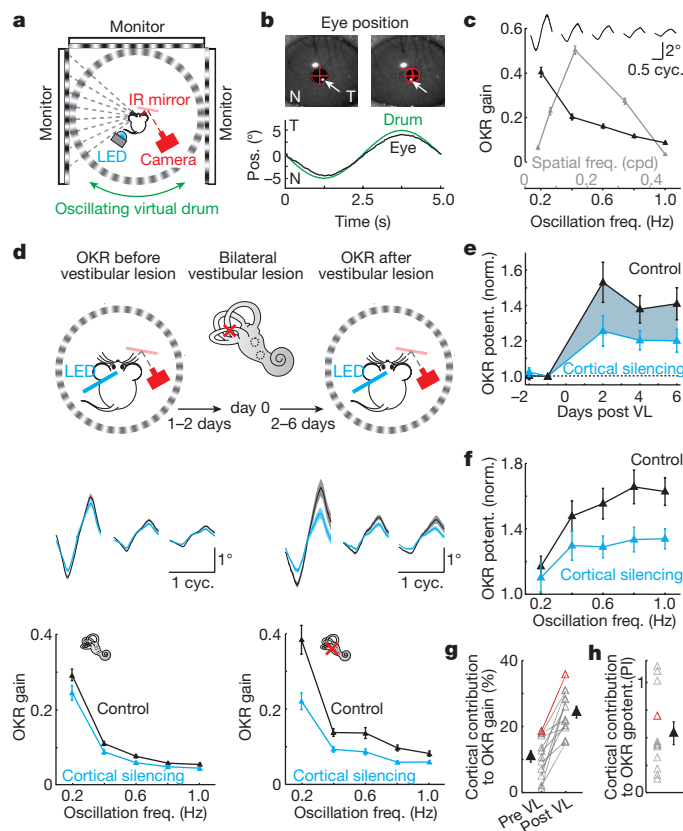


Figure 1 | Visual cortex contributes to OKR potentiation. **a**, Schematic of experimental setup. **b**, Top, snapshots of nasal (N; left) and temporal (T; right) eye positions. Red ellipses, pupil fit. Red crosses, pupil centres. Arrows, corneal reflection of reference LED. Bottom, cycle average of individual eye trajectory overlaid with drum trajectory. **c**, Data from example mouse. Oscillation frequency and spatial frequency tuning curves ($n = 32$ and 48 trials, respectively). Top traces, cycle-averaged OKR trajectories ($n = 96$ –720 cycles). Thickness, s.e.m. **d**, Data from example mouse. Top, schematic of experimental design. Bottom, oscillation frequency tuning curves of OKR gain before vestibular lesion (VL; left, $n = 60$ trials) and 2 days after vestibular lesion (right, $n = 30$ trials). Blue and black curves, with and without cortical silencing. Top traces, cycle-averaged OKR trajectories ($n = 120$ –600 cycles). Thickness, s.e.m. **e**, Population average of OKR potentiation following vestibular lesion. Black, no cortical silencing. Blue, cortical silencing ($n = 13$ mice). Shaded area illustrates cortical contribution to OKR potentiation. **f**, Population tuning curves of OKR potentiation following vestibular lesion ($n = 13$ mice). **g**, Population summary of cortical contribution to OKR gain before (Pre VL) and after vestibular lesion (Post VL). **h**, Population summary of cortical contribution to OKR potentiation (potentiation index, PI). Red data point in **g** and **h**: animal in **d**. $n = 12$ mice for **g** and **h**. Data in **c–h** shown as mean \pm s.e.m.

¹Howard Hughes Medical Institute and Center for Neural Circuits and Behavior, University of California San Diego, La Jolla, California 92093-0634, USA. ²Neurobiology Section and Department of Neuroscience, University of California San Diego, La Jolla, California 92093-0634, USA. ³Department of Neurobiology, Stanford University School of Medicine, Palo Alto, California 94305, USA. ⁴Department of Physiology, University of California San Francisco, San Francisco, California 94143-0444, USA.

A bilateral vestibular lesion (Methods) led to a robust compensatory increase in OKR gain (OKR potentiation). Mice recovered from the surgery in their home cages and the OKR was assessed before and two, four and six days after surgery (Fig. 1d, top). Two days after surgery, OKR gain was potentiated by more than 50% ($53 \pm 11\%$ (mean \pm s.e.m., $P = 4 \times 10^{-4}$; averaged over all oscillation frequencies), and this potentiation lasted for at least six days after surgery (day 4: $38 \pm 8\%$, $P = 3 \times 10^{-4}$; day 6: $41 \pm 9\%$, $P = 6 \times 10^{-4}$; Fig. 1d, e). On average, the potentiation was greater for higher oscillation frequencies (Fig. 1f). Thus, a vestibular lesion leads to strong OKR potentiation.

To investigate whether the visual cortex contributes to OKR potentiation, we silenced this area bilaterally by photostimulating cortical γ -aminobutyric acid (GABA)-releasing inhibitory neurons expressing the light-sensitive cation channel channelrhodopsin 2 (ChR2; Extended Data Fig. 2a; Methods). Before vestibular lesioning, cortical silencing (Extended Data Fig. 2b) led to only a small reduction in OKR gain (referred to as the cortical contribution to OKR gain; $10.9 \pm 1.7\%$, $P = 4 \times 10^{-5}$; averaged over all frequencies; Fig. 1d, g and Extended Data Fig. 3a–c). Notably, in the same animals, cortical silencing two days after surgery led to a much stronger reduction in OKR gain ($24.4 \pm 1.8\%$, $P = 4 \times 10^{-5}$; Fig. 1d, g and Extended Data Fig. 3a–c). Thus the cortical contribution to OKR gain more than doubles after a bilateral vestibular lesion. To quantify the contribution of the visual cortex to OKR potentiation we used a potentiation index, PI (Methods; Extended Data Fig. 3c). PI is 1 if OKR potentiation is abolished upon cortical silencing and is 0 if the cortical contribution to OKR gain is unaffected by the vestibular lesion. The cortical contribution to OKR potentiation was between 0.1 and 1.16 and averaged 0.54 ± 0.10 across all oscillation frequencies (Fig. 1h and Extended Data Fig. 3d) indicating that at least half of OKR potentiation may depend on the visual cortex. In sham-lesioned animals (see Methods), OKR potentiation was absent and the cortical contribution to OKR gain was unaffected (Extended Data Fig. 3e, f). Thus, the contribution of the visual cortex to OKR gain increases strongly after vestibular lesioning, indicating that the visual cortex has a prominent role in OKR potentiation.

Continuous OKR stimulation can also lead to OKR potentiation^{7,10}. To determine whether this less invasive form of OKR potentiation

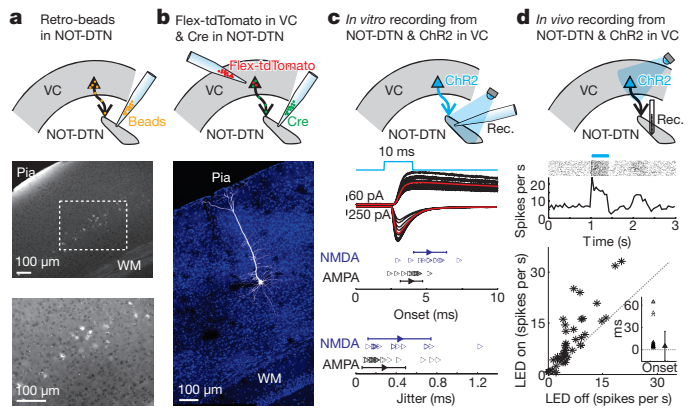


Figure 2 | Cortico-fugal projection from mouse visual cortex to NOT-DTN. **a–d**, Top, schematic of experimental design. VC, visual cortex. **a**, Beads injected into NOT-DTN are retrogradely transported to visual cortex. Middle, coronal slice of visual cortex. Bottom, higher magnification of the enclosed area. WM, white matter. **b**, Example of layer 5 pyramidal cell projecting to NOT-DTN in visual cortex. Blue, DAPI; white, tdTomato. **c**, Middle, AMPA receptor-mediated (downwards) or NMDA receptor-mediated (upwards) EPSCs evoked in NOT-DTN neurons by optogenetic stimulation of cortico-fugal axons *in vitro*. Black, individual traces; red, average traces; blue, time course of blue light illumination. Bottom, summary of onset latency and trial-by-trial jitter of EPSCs. Data shown as mean \pm s.d.; $n = 9$ mice. **d**, Firing of NOT-DTN neurons upon optogenetic activation of visual cortex *in vivo*. Middle, raster plot and PSTH of a NOT-DTN unit. Blue bar, LED illumination. Bottom, firing rates of NOT-DTN units in LED-off trials vs LED-on trials. Dotted line, unity line. Inset, summary of onset latency ($n = 23$ units). Data shown as median \pm s.d.; $n = 4$ mice.

also depends on the visual cortex, we exposed animals to continuous OKR stimulation for about 30 min. This protocol led to a progressive increase in OKR gain ($50 \pm 6\%$, $P = 4 \times 10^{-6}$; Extended Data Fig. 4a–c; Methods) and to an increase in the cortical contribution to OKR gain (from $5 \pm 2\%$ to $16 \pm 2\%$, $P = 6 \times 10^{-4}$; Extended Data Fig. 4d), accounting for $33 \pm 7\%$ of OKR potentiation ($P = 5 \times 10^{-4}$,

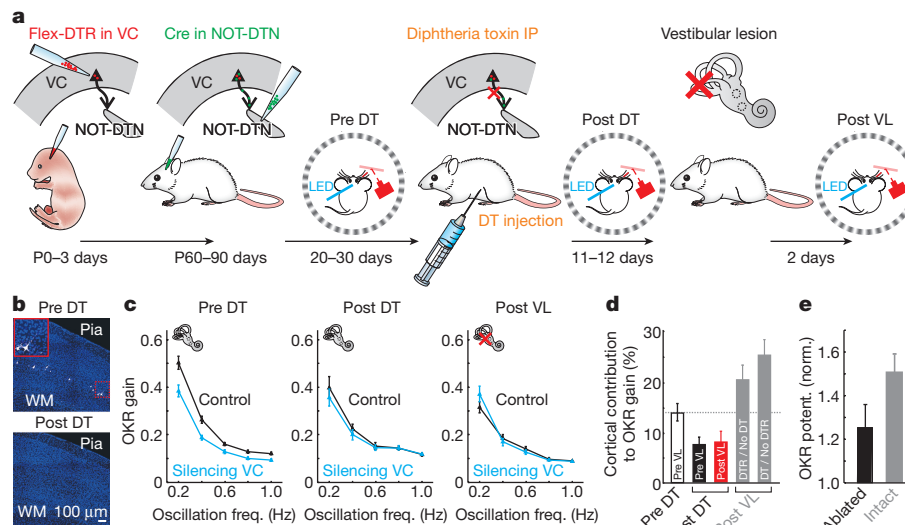


Figure 3 | Cortico-fugal projection to NOT-DTN is necessary for OKR potentiation. **a**, Schematic of experimental design. DTR, diphtheria toxin receptor; DT, diphtheria toxin; VL, vestibular lesion. **b**, Coronal slices of visual cortex from a mouse without diphtheria toxin injection (top) or with diphtheria toxin injection (bottom). Inset, higher magnification of the enclosed area. Blue, DAPI; white, GFP. **c**, Data from example mouse. Oscillation frequency tuning curves measured before ablation of the cortico-fugal projections (Pre DT, $n = 60$ trials), after ablation (Post DT, $n = 30$ trials) and after vestibular lesion (Post VL, $n = 42$ trials).

d, Population average of cortical contribution to OKR gain. White, intact cortico-fugal projection. Black, after ablation of the cortico-fugal projection and before vestibular lesion. Red, after vestibular lesion. Same set of animals ($n = 18$ mice). Grey columns, two control groups after vestibular lesion: left, without diphtheria toxin injection ($n = 6$ mice); right, without DTR infection ($n = 8$ mice). **e**, Population average of OKR potentiation for mice with intact or ablated cortico-fugal projections ($n = 27$ and 18 mice, respectively). Data in c–e shown as mean \pm s.e.m.

Extended Data Fig. 4c, e). Thus, the visual cortex contributes to OKR potentiation induced by vestibular lesioning or continuous OKR stimulation.

Through what pathway does the visual cortex influence OKR potentiation? Because the midbrain nuclei of the accessory optic system (AOS) represent the first stage downstream of the retina in the circuit that mediates the OKR^{3,5,6}, we investigated whether mouse visual cortex targets these structures. The horizontal OKR is mediated by the AOS structure composed of the apposed optic tract and dorsal-terminal nuclei (NOT-DTN)^{3,5,6}. To identify the NOT-DTN, we used the *Hoxd10*-GFP mouse, in which retinal ganglion cells (RGCs) innervating the AOS express green fluorescent protein (GFP; Extended Data Fig. 5a). We verified the extent to which *Hoxd10*-GFP fluorescence delineated the NOT-DTN using c-Fos immunostaining (Methods). OKR stimulation for 60 min enhanced c-Fos expression in neurons located at coordinates corresponding to the NOT-DTN (Extended Data Fig. 5a–d). Furthermore, *Hoxd10*-GFP fluorescence overlapped with the enhanced expression of c-Fos in the NOT-DTN (overlap coefficient $86.4 \pm 0.8\%$, $n = 43$ slices from 4 mice; Extended Data Fig. 5c). OKR stimulation did not enhance c-Fos expression in other visual nuclei not directly involved in OKR, such as the superior colliculus and ventral lateral geniculate nucleus (vLGN) (Extended Data Fig. 5e, f). Thus, we can unequivocally identify the NOT-DTN in mice. Stereotactic injections of retrograde fluorescent microspheres into the NOT-DTN (Methods) labelled layer 5 in the visual cortex (Fig. 2a) along with additional brain areas presynaptic to NOT-DTN (Extended Data Fig. 6a), indicating that the mouse visual cortex, like those of primates and carnivores^{21,22}, projects directly to NOT-DTN (Extended Data Fig. 6b).

To reveal the morphology of visual cortical neurons that projected to NOT-DTN, we injected the retrograde CAV2-Cre virus in the NOT-DTN of mice conditionally expressing tdTomato in the visual cortex (Methods). These injections revealed a sparse population of layer 5 pyramidal neurons (0.21% of layer 5 neurons) distributed across the primary and secondary visual cortices (Fig. 2b and Extended Data Fig. 6c). To verify that these cortico-fugal neurons form functional synaptic contacts with NOT-DTN neurons, we performed whole-cell recordings from NOT-DTN neurons in acute slices from mice expressing Chr2 in the visual cortex. Photo-stimulation of cortical axons triggered excitatory postsynaptic currents in NOT-DTN neurons (166 ± 258 pA, mean \pm s.d.) with a short latency (3.9 ± 0.8 ms, mean \pm s.d.) and little jitter (0.28 ± 0.21 ms, mean \pm s.d.), mediated by both AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) and NMDA (*N*-methyl-D-aspartate) receptors (Fig. 2c and Extended Data Fig. 6d). The monosynaptic identity of this projection was validated using subcellular Chr2-assisted circuit mapping (sCRACM; Extended Data Fig. 6e). Finally, to determine whether this cortico-fugal projection could drive NOT-DTN neurons, we recorded extracellularly from the NOT-DTN in anaesthetized mice (Methods). Optogenetic activation of the visual cortex increased the firing of NOT-DTN neurons with a short latency (5 ± 19.7 ms, median \pm s.d.; Fig. 2d). Thus, cortico-fugal projections from mouse visual cortex to the NOT-DTN form functional excitatory synapses that can drive the activity of NOT-DTN neurons.

To determine whether cortico-fugal projections to the NOT-DTN are necessary for the cortical component of OKR potentiation, we selectively ablated NOT-DTN-projecting cortical neurons with diphtheria toxin. We infected the NOT-DTN with the retrograde CAV2-Cre virus in mice whose visual cortex had been injected with an AAV virus expressing the diphtheria toxin receptor (DTR) in a Cre-dependent manner (Methods; Fig. 3a). DTR-expressing neurons in the visual cortex were completely ablated 11 days after intraperitoneal diphtheria toxin injection (Fig. 3b). We tested the OKR and its modulation by visual cortex in the same mice at three time points: before diphtheria toxin injection, after diphtheria toxin injection and after vestibular lesioning (Fig. 3a, c). Before diphtheria toxin injection, cortical silencing reduced OKR gain by $14 \pm 2\%$ ($P = 4 \times 10^{-7}$; Fig. 3d), similar

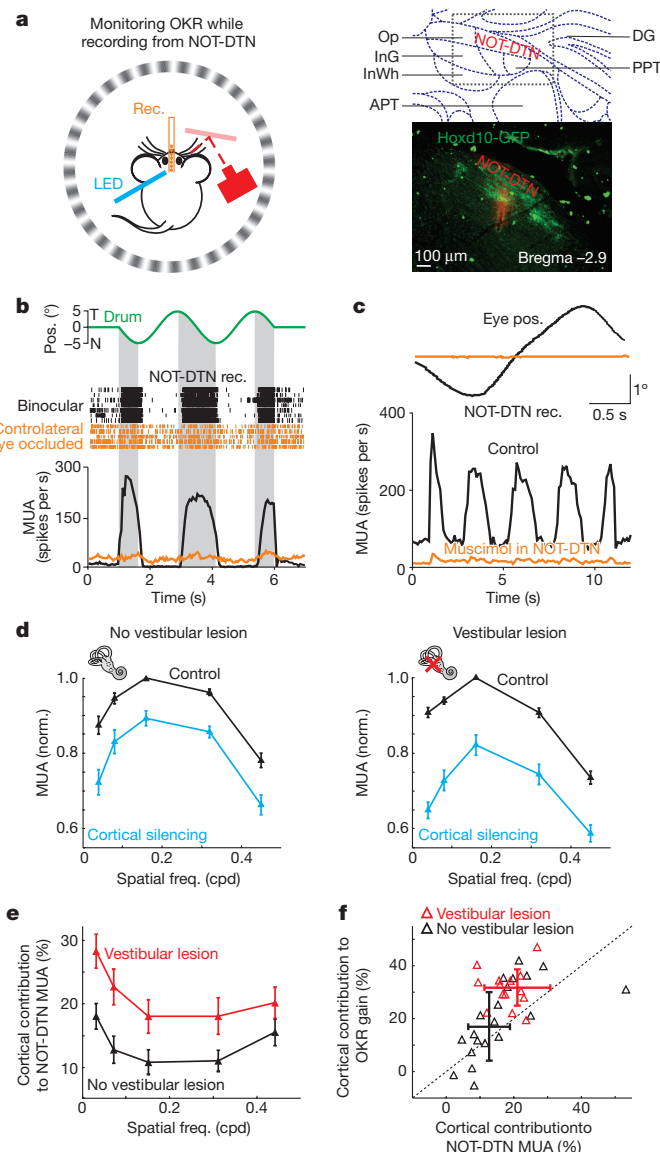


Figure 4 | Enhanced cortical modulation of NOT-DTN activity with OKR potentiation. **a**, Left, schematic of experimental setup. Right, top, NOT-DTN and surrounding nuclei (modified from Paxinos, G. & Franklin, K. *The Mouse Brain in Stereotaxic Coordinates* (Elsevier, 2007)); bottom, coronal slice of NOT-DTN from *Hoxd10*-GFP mouse. Green, GFP; red, electrode track labelled with DiI. **b**, Data from example mouse. Top, drum trajectory. Shaded areas, temporo-nasal phase of drum trajectory. Bottom, raster plot and post-stimulus time histogram (PSTH) of NOT-DTN multiunit activity ($n = 30$ and 18 trials for binocular (black) and ipsilateral vision (orange), respectively). **c**, Data from example mouse. Cycle averages of all eye trajectories (top) and simultaneously recorded NOT-DTN activity (bottom) before (black; $n = 49$ trials) and after silencing NOT-DTN with muscimol (orange; $n = 49$ trials). **d**, Population tuning curves of NOT-DTN activity for naive mice (no vestibular lesion, $n = 17$ mice) and mice with vestibular lesion ($n = 17$ mice). Curves normalized to best frequency in control. Data shown as mean \pm s.e.m. **e**, Population tuning curves of cortical contribution to NOT-DTN activity ($n = 17$ mice for both groups). Data shown as mean \pm s.e.m. **f**, Correlation between cortical contribution to OKR gain and cortical contribution to NOT-DTN activity in naive animals (black, no vestibular lesion) and lesioned animals (red; correlation coefficient, 0.55). Data shown as mean \pm s.d.; dotted line, unity line.

to control mice (Fig. 1g). By contrast, 11–12 days after diphtheria toxin injection, cortical silencing reduced OKR gain significantly less than before injection ($7 \pm 2\%$, $P = 0.007$; Fig. 3d), demonstrating that NOT-DTN-projecting cortical neurons modulate OKR gain.

Crucially, two days after vestibular lesioning, OKR potentiation was significantly reduced in animals with an ablated cortico-NOT-DTN projection (1.25 ± 0.10 ; Fig. 3e) as compared to animals with an intact projection (that is, in animals injected with AAV-DTR and CAV-Cre but not diphtheria toxin (1.46 ± 0.13 ; $n = 6$)) or in control animals (1.51 ± 0.08 ; Fig. 3e). Furthermore, the residual OKR potentiation was nearly independent of visual cortex because cortical silencing led to a very small reduction in OKR gain ($8 \pm 2\%$; Fig. 3d and Extended Data Fig. 7), similar to the reduction observed in the same animals after diphtheria toxin injection but before vestibular lesioning. In these animals, the cortical contribution to OKR potentiation was only -0.02 ± 0.13 , significantly smaller than in control animals (0.49 ± 0.11 ; $P = 0.003$). Thus, despite its sparseness, the population of visual cortico-fugal neurons projecting to the NOT-DTN is necessary for a large fraction of OKR potentiation and is entirely responsible for the cortical contribution to this phenomenon.

Because the cortical component of OKR potentiation relies on cortico-fugal projections to the NOT-DTN, we tested whether vestibular lesioning enhanced the ability of these projections to drive NOT-DTN activity. We targeted the NOT-DTN with extracellular linear probes as above (Fig. 4a). Isolated NOT-DTN units showed a preference for temporonally moving visual stimuli presented to the contralateral eye, consistent with recordings in other mammals^{23,24} (Fig. 4b and Extended Data Fig. 8a–e). Furthermore, local application of the GABA_A-receptor agonist muscimol suppressed NOT-DTN activity ($57 \pm 7\%$) and eliminated the OKR ($90 \pm 2\%$), consistent with the necessity of this structure to trigger the reflex²⁵ (Fig. 4c and Extended Data Fig. 8f). Optogenetic silencing of visual cortex led to a stronger reduction in NOT-DTN activity in mice that had undergone vestibular lesioning than in naive mice ($21.1 \pm 2.3\%$ versus $12.7 \pm 1.5\%$, $P = 0.003$; Fig. 4d, e). Furthermore, the reduction in NOT-DTN activity correlated well with the simultaneously observed reduction in OKR gain (correlation coefficient 0.55, $P = 0.0007$; Fig. 4f). Thus, after a

vestibular lesion, NOT-DTN activity depends more on visual cortex than in naive animals.

Can the cortical contribution to OKR potentiation be accounted for by the influence of the cortex on NOT-DTN activity? NOT-DTN-projecting cortical neurons could also contribute to OKR potentiation through collateral projections to targets downstream from the NOT-DTN. We addressed this question functionally and anatomically. We first established the ‘transfer function’—the relationship between OKR gain and NOT-DTN activity (Fig. 5a). If the cortical contribution to OKR potentiation is mediated through its impact on NOT-DTN activity, reducing NOT-DTN activity through cortical silencing will lead to a reduction in OKR gain predicted by the transfer function (Fig. 5a, left). If, alternatively, visual cortex also mediates OKR potentiation through projections downstream from the NOT-DTN, cortical silencing will lead to a decrease in OKR gain not predicted by the transfer function (Fig. 5a, right). We recorded NOT-DTN activity while monitoring the OKR in response to stimuli of various spatial frequencies that elicit a wide range of OKR velocities (Extended Data Fig. 9a). The dependence of NOT-DTN activity on the spatial frequency of stimuli was similar to that of the OKR gain (Extended Data Fig. 9b, c and Fig. 1c; this was not the case for other visual nuclei, Extended Data Fig. 9e, f). We fitted the transfer function $G = kR^x$ to the data points relating NOT-DTN activity (R) to OKR gain (G ; x , exponent; k , proportionality factor) (Fig. 5b and Extended Data Fig. 9d). In both naive and lesioned animals, the reduction in OKR gain that resulted from cortical silencing was accurately predicted by the concomitant reduction in NOT-DTN activity ($P = 0.71$ for naive, 0.84 for lesion, Kolmogorov–Smirnov test; Fig. 5c–e). Cortical silencing simply shifted the data points along the transfer function obtained without cortical silencing, bringing them closer to the origin. This shift was much larger in lesioned than in naive animals (0.19 ± 0.01 versus 0.09 ± 0.01 , lesioned versus naive; $P = 7 \times 10^{-10}$; quantified as vector length, see Methods; Extended Data Fig. 9g–j). Thus, the cortical contribution to

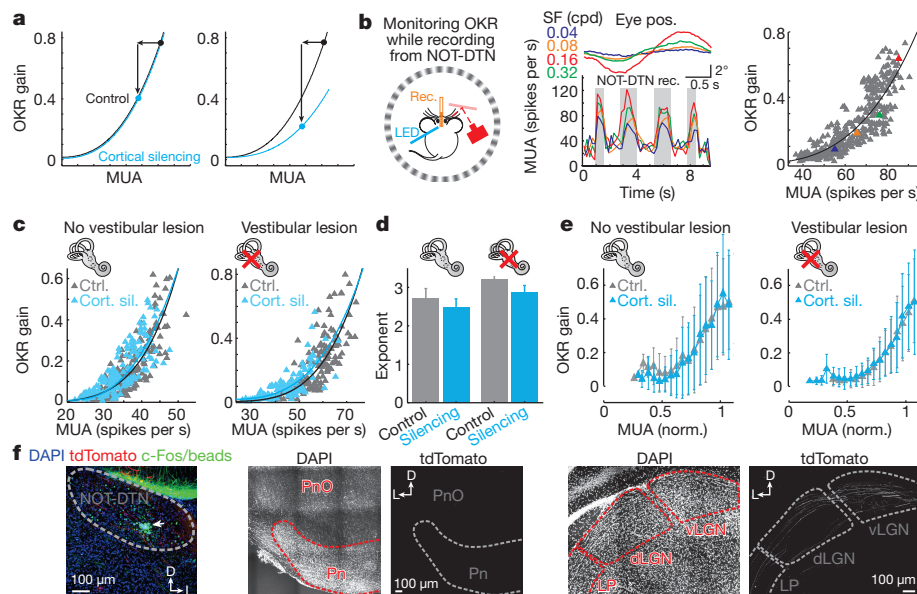


Figure 5 | Impact of cortex on NOT-DTN activity matches cortical contribution to OKR potentiation. **a**, Model transfer functions. Left, reduction in NOT-DTN activity (leftward arrow) upon cortical silencing leads to a reduction in OKR gain (downward arrow) predicted by the transfer function obtained under control conditions. Right, reduction in OKR gain upon cortical silencing is larger than predicted by the transfer function. **b**, Data from example mouse. Left, schematic of experimental setup. Middle, cycle averages of OKR trajectory (top, $n = 3$ cycles) and PSTH of simultaneously recorded NOT-DTN activity (bottom) in individual trials of four different spatial frequencies (colour coded). Right, transfer function. Each data point is one trial. Coloured triangles,

trials illustrated in the middle. MUA, firing rate during the temporo-nasal phase (shaded in PSTH). Solid curve, best fit of power function. **c**, Data from example mice. Transfer functions of a naive mouse (no vestibular lesion) and a mouse with vestibular lesion. **d**, Population averages of exponent of transfer functions for naive (non-lesioned, $n = 17$) and lesioned mice ($n = 17$). **e**, Population-averaged normalized transfer functions for naive (non-lesioned) and lesioned mice. Data shown as mean \pm s.d. **f**, Left, confocal image of NOT-DTN. White arrow, injection site of CAV2-Cre virus. Middle and right, confocal images of pontine nuclei (Pn) and visual thalamus (LP, dLGN and vLGN). D, dorsal; L, lateral.

OKR potentiation can be fully accounted for by its increased impact on NOT-DTN activity. Finally, we determined whether NOT-DTN-projecting cortical neurons send collaterals to additional subcortical structures. Consistent with the above functional results, we did not observe collaterals of NOT-DTN-projecting cortical neurons in downstream structures involved in OKR (the pontine nuclei²⁶, pre-oculomotor nuclei near the periaqueductal grey, inferior olive or cerebellum; Fig. 5f and Extended Data Fig. 10). Instead, we observed labelled axons in the vLGN, superior colliculus and striatum. As the vLGN and superior colliculus project to the NOT-DTN^{3,27} (Extended Data Fig. 6a), they could also contribute to the increased drive of the NOT-DTN. However, OKR stimulation evoked only weak activity in the superior colliculus or vLGN (multiunit activity, 2.2 ± 0.4 spikes per s for superior colliculus and 1.2 ± 0.3 spikes per s for vLGN) and this activity did not correlate with OKR gain (Extended Data Fig. 9e, f). These two structures are thus unlikely to contribute substantially to the cortical component of OKR potentiation (see also Extended Data Fig. 5e, f).

These results show that the visual cortex has an essential role in OKR potentiation and identify the cortico-fugal projection to the NOT-DTN as the anatomical and physiological substrate that underlies OKR potentiation.

The cortex-independent fraction of OKR potentiation may be mediated by subcortical structures such as the cerebellum and vestibular nuclei, consistent with their established roles in OKR plasticity^{10–13}. Thus, cortico-fugal projections from the visual cortex to the brainstem may work with these subcortical structures to mediate the compensatory potentiation of the reflex.

Cortico-fugal projections from sensory areas to brainstem nuclei can modulate innate behaviours^{28,29} and learning-dependent plasticity³⁰, but our understanding of the functions of those projections is still very limited. In primates and carnivores, the visual cortex contributes to some properties of the OKR, including directional symmetry of the reflex and gain⁵. The demonstration that cortico-fugal projections to the AOS play an essential role in the plastic adaptation of the OKR expands our understanding of how the innervation of phylogenetically older structures by the mammalian cortex can improve the performance of reflexive behaviour in an experience-dependent manner. Thus, the visual cortex must be regarded not only as an area for sensory processing but as an area that, through its output to the brainstem, is directly involved in the plasticity of fundamental innate motor behaviours.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 March; accepted 2 September 2016.

Published online 12 October 2016.

- Sherman, S. M. & Guillery, R. W. *Functional Connections of Cortical Areas: A New View from the Thalamus* (MIT Press, 2013).
- Guillery, R. W. Anatomical pathways that link perception and action. *Prog. Brain Res.* **149**, 235–256 (2005).
- Giolli, R. A., Blanks, R. H. & Lui, F. The accessory optic system: basic organization with an update on connectivity, neurochemistry, and function. *Prog. Brain Res.* **151**, 407–440 (2006).
- Kawai, R. et al. Motor cortex is required for learning but not for executing a motor skill. *Neuron* **86**, 800–812 (2015).
- Distler, C. & Hoffmann, K. P. *The Oxford Handbook of Eye Movements* 65–83 (Oxford Univ. Press, 2011).
- Simpson, J. I. The accessory optic system. *Annu. Rev. Neurosci.* **7**, 13–41 (1984).
- Faulstich, B. M., Onori, K. A. & du Lac, S. Comparison of plasticity and development of mouse optokinetic and vestibulo-ocular reflexes suggests differential gain control mechanisms. *Vision Res.* **44**, 3419–3427 (2004).
- Lisberger, S. G., Miles, F. A., Optican, L. M. & Eighmy, B. B. Optokinetic response in monkey: underlying mechanisms and their sensitivity to long-term adaptive changes in vestibuloocular reflex. *J. Neurophysiol.* **45**, 869–890 (1981).

- Prusky, G. T., Silver, B. D., Tschetter, W. W., Alam, N. M. & Douglas, R. M. Experience-dependent plasticity from eye opening enables lasting, visual cortex-dependent enhancement of motion vision. *J. Neurosci.* **28**, 9817–9827 (2008).
- Katoh, A., Kitazawa, H., Itoharu, S. & Nagao, S. Dynamic characteristics and adaptability of mouse vestibulo-ocular and optokinetic response eye movements and the role of the flocculo-olivary system revealed by chemical lesions. *Proc. Natl Acad. Sci. USA* **95**, 7705–7710 (1998).
- Faulstich, M., van Alphen, A. M., Luo, C., du Lac, S. & De Zeeuw, C. I. Oculomotor plasticity during vestibular compensation does not depend on cerebellar LTD. *J. Neurophysiol.* **96**, 1187–1195 (2006).
- Gittis, A. H. & du Lac, S. Intrinsic and synaptic plasticity in the vestibular system. *Curr. Opin. Neurobiol.* **16**, 385–390 (2006).
- Kawato, M. & Gomi, H. The cerebellum and VOR/OKR learning models. *Trends Neurosci.* **15**, 445–453 (1992).
- Fetter, M., Zee, D. S. & Proctor, L. R. Effect of lack of vision and of occipital lobectomy upon recovery from unilateral labyrinthectomy in rhesus monkey. *J. Neurophysiol.* **59**, 394–407 (1988).
- Tusa, R. J., Demer, J. L. & Herdman, S. J. Cortical areas involved in OKN and VOR in cats: cortical lesions. *J. Neurosci.* **9**, 1163–1178 (1989).
- Fetter, M. & Zee, D. S. Recovery from unilateral labyrinthectomy in rhesus monkey. *J. Neurophysiol.* **59**, 370–393 (1988).
- Paige, G. D. Plasticity in the vestibulo-ocular and optokinetic reflexes following modification of canal input. *Rev. Oculomot. Res.* **1**, 145–153 (1985).
- McCall, A. A. & Yates, B. J. Compensation following bilateral vestibular damage. *Front. Neurol.* **2**, 88 (2011).
- Reid, R. C. & Alonso, J. M. The processing and encoding of information in the visual cortex. *Curr. Opin. Neurobiol.* **6**, 475–480 (1996).
- Tabata, H., Shimizu, N., Wada, Y., Miura, K. & Kawano, K. Initiation of the optokinetic response (OKR) in mice. *J. Vision* **10**, 11–17 (2010).
- Distler, C. & Hoffmann, K. P. Cortical input to the nucleus of the optic tract and dorsal terminal nucleus (NOT-DTN) in macaques: a retrograde tracing study. *Cereb. Cortex* **11**, 572–580 (2001).
- Berson, D. M. & Graybiel, A. M. Some cortical and subcortical fiber projections to the accessory optic nuclei in the cat. *Neuroscience* **5**, 2203–2217 (1980).
- Hoffmann, K. P. & Distler, C. Quantitative analysis of visual receptive fields of neurons in nucleus of the optic tract and dorsal terminal nucleus of the accessory optic tract in macaque monkey. *J. Neurophysiol.* **62**, 416–428 (1989).
- Grasse, K. L. & Cynader, M. S. Electrophysiology of lateral and dorsal terminal nuclei of the cat accessory optic system. *J. Neurophysiol.* **51**, 276–293 (1984).
- Collewin, H. Oculomotor areas in the rabbits midbrain and pretectum. *J. Neurobiol.* **6**, 3–22 (1975).
- Distler, C., Mustari, M. J. & Hoffmann, K. P. Cortical projections to the nucleus of the optic tract and dorsal terminal nucleus and to the dorsolateral pontine nucleus in macaques: a dual retrograde tracing study. *J. Comp. Neurol.* **444**, 144–158 (2002).
- Baldauf, Z. B., Wang, X. P., Wang, S. & Bickford, M. E. Pretectotectal pathway: an ultrastructural quantitative analysis in cats. *J. Comp. Neurol.* **464**, 141–158 (2003).
- Xiong, X. R. et al. Auditory cortex controls sound-driven innate defense behaviour through corticofugal projections to inferior colliculus. *Nat. Commun.* **6**, 7224 (2015).
- Liang, F. et al. Sensory cortical control of a visually induced arrest behavior via corticotectal projections. *Neuron* **86**, 755–767 (2015).
- Bajo, V. M., Nodal, F. R., Moore, D. R. & King, A. J. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* **13**, 253–260 (2010).

Acknowledgements We thank members of the Scanziani and Isaacson laboratories for advice on this project; J. Evora, N. Kim, M. Chan and A. Linder for technical support; T. M. Jessell for sharing the Flex.DTR.GFP plasmid and virus; M. Faulstich and S. du Lac for sharing eye tracking codes and advice on vestibular lesions; D. Li for advice on statistical analysis; J. Isaacson for comments on the manuscript; S. R. Olsen for sharing codes for *in vivo* recordings and unit isolation; and M. Xue for help with *in vitro* recordings. M.S. is an investigator of the Howard Hughes Medical Institute. This work was supported by the Gatsby Charitable Foundation and the US National Institutes of Health (R01 EY025668).

Author Contributions B.L. and M.S. designed the study. B.L. performed all experiments and data analysis. A.D.H. shared the Hoxd10–GFP mouse line. B.L. and M.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.L. (lbaohua@ucsf.edu) or M.S. (massimo@ucsf.edu).

METHODS

Mice. Experiments were performed in accordance with the regulations of the Institutional Animal Care and Use Committee of the University of California, San Diego.

We used the following mouse lines: VGAT-ChR2-EYFP³¹ (Jackson Labs #014548), PV-Cre³² (Jackson Labs #008069), Gad2-Cre³³ (Jackson Labs #010802) and Hoxd10-GFP³⁴ (MMRRC #032065-UCD). Mice were bred by crossing homozygous VGAT-ChR2-EYFP, PV-Cre or Gad2-Cre males (all lines with a C57BL/6 background) with wild-type ICR females or homozygous Hoxd10-GFP females (ICR background) to C57BL/6 males. Mice were housed in a vivarium with a reversed light cycle (12 h day–12 h night). Mice of both genders were used for experiments at postnatal ages of 2–6 months.

Viral and retrobead injections. We used the following adeno-associated viruses (AAV) and canine adenovirus (CAV2):

For the Cre recombinase (Cre)-dependent expression of Channelrhodopsin2 (ChR2)^{35,36}, AAV2/9.CAGGS.Flex.ChR2.tdTomato.SV40 (Addgene 18917; UPenn Vector Core).

For the Cre-dependent expression of tdTomato: AAV2/1.CAG.Flex.tdTomato.WPRE.bGH (Allen Institute 864; UPenn Vector Core).

For the expression of Cre: AAV2/9.hSyn.HI.eGFP-Cre.WPRE.SV40 (UPenn Vector Core).

For Cre-dependent expression of the diphtheria toxin receptor (DTR)³⁷: AAV2/1.Flex.DTR.GFP (Jessell laboratory; produced at UNC Vector Core).

For retrograde expression of Cre: CAV2.Cre³⁸ (Montpellier vector platform).

Optogenetic silencing of visual cortex. AAV2/9.CAGGS.Flex.ChR2.tdTomato.SV40 was bilaterally injected into the visual cortex of newborn PV-Cre or Gad2-Cre pups (postnatal day (P) 0–2). The virus was loaded into a bevelled glass micropipette (tip diameter 20–40 µm) mounted on a Nanoject II (Drummond) attached to a micromanipulator. Pups were anaesthetized by hypothermia and secured in a molded platform. In each hemisphere the virus was injected at two sites along the medial–lateral axis of the visual cortex. At each site we made three bolus injections of 28 nl. Each were at three different depths between 300 and 600 µm. Protein expression was verified by epi-fluorescent illumination through a dissection microscope (Leica MZ10F). Experiments were performed on animals with expression over the entire extent of visual cortex.

Optogenetic stimulation of visual cortex *in vivo* or cortico-fugal axons *in vitro*. AAV2/9.hSyn.HI.eGFP-Cre.WPRE.SV40 and AAV2/9.CAGGS.Flex.ChR2.tdTomato were mixed in 1:20 ratio. The mixture was injected into the visual cortex of newborn C57BL/6 pups (as described above). Protein expression was verified by epi-fluorescent illumination.

Retrogradely labelling of NOT-DTN-projecting neurons in the visual cortex. Adult Hoxd10-GFP mice were anaesthetized with ~2% isoflurane (vol/vol) in O₂. The depth of anaesthesia was monitored with the toe-pinch response. The eyes were protected from drying by artificial tears. We cut open the scalp and thinned the skull to create a window of ~300–500 µm diameter. The remaining layer of bone in the window was thin enough to allow the penetration of the bevelled glass pipette. A bolus of retrograde fluorescent microspheres (RetroBeads, Lumafuor Inc.) or CAV2.Cre virus (40 nl RetroBeads or 20 nl CAV2 virus) was injected into the NOT-DTN (coordinates (anteroposterior axis (AP) relative to bregma; mediolateral axis (ML) relative to the midline): AP: –1,260 µm; ML: 3,080 µm; depth: 1,960 µm; coordinates were adjusted based on the distance between bregma and lambda on mouse skull) using an UltraMicroPump (UMP3, WPI). The wound was sutured with a few stitches of 6-0 suture silk (Fisher Scientific NC9134710). Mice were perfused 3 days after the retrobead injection or 2 weeks after the CAV2 injection.

Ablation of the cortico-fugal projection to the NOT-DTN. AAV2/1.Flex.DTR.GFP was bilaterally injected into the visual cortex of VGAT-ChR2-EYFP pups between P0 and P2. CAV2.Cre virus was subsequently stereotactically injected into the NOT-DTN (same coordinates as above) bilaterally in mice of 2–6 months of age. Three to four weeks later we injected diphtheria toxin (DT 40 ng/g) intraperitoneally three times on alternate days. The OKR was assessed 11 or 12 days after the first diphtheria toxin injection. In control experiments, diphtheria toxin was replaced with PBS or diphtheria toxin was injected into mice that had not been infected with AAV2/1.Flex.DTR.GFP.

Head bar implantation and cranial window. Mice were implanted with a T-shaped head bar for head fixation. Mice were anaesthetized using ~2% isoflurane. The scalp and fascia were removed and a metal head bar was mounted over the midline using dental cement (Ortho-Jet powder; Lang Dental) mixed with black paint (iron oxide).

We created a cranial window of ~3 × 3 mm (1.5–4.5 mm lateral to midline and 2.3–5.2 mm posterior to bregma) over the visual cortex on each hemisphere by gently thinning the skull until it appeared transparent when wetted by

saline solution. The window was then covered with a thin layer of crazy glue. Following the surgery animals were injected subcutaneously with 0.1 mg/kg buprenorphine and allowed to recover in their home cage for at least 1 week. Several days before the test, mice were familiarized with head fixation in the recording setup. No visual stimulation was given.

Assessment of the OKR. Visual stimulation. The horizontal OKR was elicited by a 'virtual drum' system³⁹. Three computer LED monitors (Viewsonic VX2450wm-LED, 60-Hz refresh rate, gamma-corrected) were mounted orthogonally to each other to form a square enclosure that covered ~270° of visual field along the azimuth. The mouse head was immobilized at the centre of the enclosure with the nasal and temporal corners of the eye leveled. Visual stimuli were generated with Psychophysics Toolbox 3 running in Matlab (Mathworks). To ensure synchronized updating across multiple monitors we used AMD Eyefinity Technology (ATI FirePro V4800). The monitors displayed a vertical sinusoidal grating whose period (spacing between stripes) was adjusted throughout the azimuthal plane such that the projection of the grating on the eye had constant spatial frequency. In other words, the spatial frequency of the grating was perceived as constant throughout the visual field, as if the grating was drifting along the surface of a virtual drum. The dependence of pixel brightness on monitor coordinates was obtained by using this equation: $B = L + L \times C \times \sin(2\pi \times x_{\text{deg}} \times \text{SF})$, where B is the brightness of pixels, L is the luminance in cd/m², C is the contrast, SF is the spatial frequency and x_{deg} is the azimuth of pixels in degrees, which is transformed from the Cartesian coordinates of the monitor into the cylindrical coordinates of the virtual drum by the following formula: $x_{\text{deg}} = \tan^{-1}(x_{\text{pix}}/D)$, where x_{pix} is the horizontal pixel position in Cartesian coordinates and D is the distance from the centre of the monitors to the eye (Extended Data Fig. 1a).

The grating drifted clockwise or counterclockwise in an oscillatory manner^{7,11} (oscillation amplitude ± 5°; grating spatial frequency: 0.04–0.45 cpd; oscillation frequency 0.2–1 Hz, corresponding to a peak velocity of the stimulus of 6.28–31.4° s^{–1}; contrast: 80%; mean luminance: 40 cd/m²). We chose the duration of the visual stimulus to allow the presentation of an integral number of oscillatory cycles (10 or 15 s for OKR test only; 7.5 s for simultaneous NOT-DTN electrophysiology and OKR test). Trials were spaced by an inter-stimulation interval of at least 8 s. The inter-stimulation interval following trials of cortical silencing was increased to 20 s. To measure the oscillation frequency tuning, spatial frequency was kept constant at 0.08 cpd; to measure the spatial frequency tuning oscillation, the frequency was kept at 0.4 Hz.

To obtain the transfer function, we varied the spatial frequency of the visual stimulus rather than the oscillation frequency because OKR peak velocity is strongly modulated by spatial frequency and much less so by the oscillation frequency (consistent with previous observations^{7,40}; Extended Data Fig. 9a). The spatial frequency was varied from 0.04 to 0.45 cpd, and the oscillation frequency was kept constant at 0.4 Hz.

To evaluate the directional preference of NOT-DTN neurons, one monitor was positioned 20 cm from the eye contralateral to the side of recording. Full-field sinusoidal drifting gratings (oscillation frequency: 1 Hz; spatial frequency: 0.08 cpd; mean luminance: 50 cd/m²; contrast: 100%) were used. Gratings were randomly presented at 12 equally spaced positions. The duration of the visual stimulus was 2 s and the inter-trial interval was 2.2 s.

To visualize NOT-DTN with c-Fos immunostaining (c-Fos is an immediate early gene expressed in response to neuronal activity), OKR was elicited by drum stimulation of various spatial frequencies (0.04–0.45 cpd) with oscillation frequency 0.4 Hz, contrast 100% and luminance 50 cd/m². Trials of oscillatory motion lasted for 15 s and were followed by an inter-trial interval of 8 s. The whole stimulation procedure took 60 min.

Monitoring eye movements by infrared video-oculography^{41,42}. The movement of the right eye was monitored through a high speed infrared (IR) camera (Imperx IPX-VGA 210; 100 Hz). The camera captured the reflection of the eye on an IR mirror (transparent to visible light, Edmund Optics #64-471) under the control of custom labview software and a frame grabber (National Instrument PCIe-1427). The pupil was identified online by thresholding pixel values or *post hoc* by combining thresholding and morphology operation and its profile was fitted with an ellipse to determine the centre. The eye position was measured by computing the distance between the pupil centre and the corneal reflection of a reference IR LED placed along the optical axis of the camera. To calibrate the measurement of the eye position, the camera and the reference IR LED were moved along a circumference centred on the image of the eye by ± 10° (Extended Data Fig. 1b).

Optogenetic silencing⁴³ or stimulation of the visual cortex. Three mouse lines (VGAT-ChR2-EYFP, PV-Cre and Gad2-Cre) were used in experiments involving optogenetic silencing of the visual cortex. They are equally efficient in silencing activity of visual cortex and interchangeable. VGAT-ChR2-EYFP mice were used in most of the silencing experiments, except in experiments illustrated

in Extended Data Fig. 2a (PV-Cre line) and Extended Data Fig. 3b (all 3 lines). To photostimulate ChR2-expressing cortical inhibitory neurons *in vivo*, a 470-nm blue fibre-coupled LED (1 mm diameter, Doric Lenses) was placed ~5–10 mm above the cranial windows of each hemisphere. We restricted the illumination to the tissue under the cranial window by covering neighbouring areas with dental cement. An opaque shield of black clay prevented LED light from directly reaching the eyes. The total light power out of the LED fibre was 15–20 mW. Trials were alternated between visual stimulus alone and visual stimulus plus LED. The LED was turned on during the whole period of visual stimulation and turned off by ramping down the power over 0.5 s to limit rebound activation of the visual cortex.

To photostimulate cortical input to the NOT-DTN *in vivo*, blue light illuminated only the visual cortex ipsilaterally to the NOT-DTN where the probe was inserted. **Vestibular lesions.** We dissected out the tissue overlying the horizontal semicircular canal in mice under ~2% isoflurane anaesthesia. A small hole was drilled in the canal with a miniature Busch Bur (0.25 mm, Gesswein) and the endolymph was partially drained. The horizontal semicircular canal was plugged with bone wax (FST 19009-00) to seal the opening and reduce the flow of the endolymph within the canal. The wound was sutured with a few stitches of 6/0 suture. Mice recovered for two days in their home cages before being tested for OKR. Sham lesions were done in the same way except that no hole was drilled and no wax was introduced in the semicircular canal.

Continuous OKR stimulation. OKR gain (spatial frequency: 0.1 cpd; oscillation frequency: 0.4 Hz; contrast: 100%; mean luminance: 35 cd/m²) was assessed 1 day before and 1 h before OKR training. Two sessions (12 min) were used to minimize the effect of visual stimulation during OKR evaluation on OKR gain. During continuous OKR stimulation, a drum of the same visual parameters ran continuously for 38 min. OKR gain was then assessed again 12 min after OKR stimulation was finished.

***In vivo* recordings from the NOT-DTN, superior colliculus or vLGN of awake or anaesthetized mice.** Mice were implanted with a T-shaped head bar for head fixation in the same way as described above for the OKR assessment, except that the procedure was done stereotactically with the help of an inclinometer (Digi-Key electronics 551-1002-1-ND). The inclinometer allowed us to calibrate the inclination of the two axis of the T bar relative to the anteroposterior (AP) and mediolateral (ML) axes of the skull before fixing it to the skull with dental cement. Three reference points with known coordinates were marked on the mouse skull because both bregma and lambda were inevitably masked by the dental cement holding the head bar. The head post on the recording rig was also calibrated with the same inclinometer to ensure that the recording probes were in register with the skull.

Recordings from awake animals were performed using a method similar to that described previously⁴³. One to two weeks before recording, mice were familiarized with head fixation within the recording setup over the course of two to four 50-min sessions. One day before recording, mice were anaesthetized with ~2% isoflurane. Whiskers and eyelashes contralateral to the recording side were trimmed to prevent interference with infrared video-oculography. To access the NOT-DTN we made an elongated, anteroposteriorly oriented craniotomy (~0.4 × 0.8 mm) around the coordinates of −3 mm (anteroposterior) and 1.3 mm (mediolateral). The coordinates were adjusted based on the distance between bregma and lambda on mouse skull. The craniotomy was then covered by Kwik-Cast Sealant (WPI).

On the day of recording, after peeling off the Kwik-Cast cover, a drop of artificial cerebrospinal fluid (ACSF; in mM, 140 NaCl, 2.5 KCl, 2.5 CaCl₂, 1.3 MgSO₄, 1.0 NaH₂PO₄, 20 HEPES and 11 glucose, pH 7.4) was placed in the well of the craniotomy to keep the exposed brain moist. A 16-channel linear silicon probe (NeuroNexus a1x16-5mm-25-177) mounted on a manipulator (Luigs & Neumann) was slowly advanced into the brain to a depth of 2,000–2,200 μm. The occurrence of direction modulated activity upon visual stimulation was used to identify the NOT-DTN (see data analysis below). The probe was stained by lipophilic DiI to label the recording track for *post hoc* verification of successful targeting of the NOT-DTN.

Recordings were not started until 20 min after insertion of the probe into the NOT-DTN. Signals were amplified 400-fold, band-pass filtered (0.3–5,000 Hz, with the presence of a notch filter) with an extracellular amplifier (A-M Systems 3600) and digitized at 32 kHz (National Instrument PCIe-6259) with custom-written software in Matlab. Raw data were stored on a computer hard drive for offline analysis. At the end of the recording session, brains were fixed by transcardial perfusion of 4% paraformaldehyde for histological analysis.

Recordings from the superior colliculus or vLGN were done in the same way except that the coordinates of the craniotomy were 3.5 mm (anteroposterior) and 1 mm (mediolateral) for the superior colliculus and 2.5 mm (anteroposterior) and 2.3 mm (mediolateral) for the vLGN.

For recordings from anaesthetized mice we used the same procedures as described above except that (1) the familiarization step was omitted and the craniotomy was performed immediately before recording; (2) animals were

anaesthetized with urethane (1.2 g/kg, intraperitoneal) and given the sedative chlorprothixene (0.05 ml of 4 mg/ml, intramuscular), as previously described⁴⁴; (3) body temperature was maintained at 37°C using a feedback-controlled heating pad (FHC 40-90-8D); (4) a uniform layer of silicone oil was applied to the eyes to prevent drying; and (5) lactated Ringer's solution was administered at 3 ml/kg/h to prevent dehydration.

***In vitro* recordings.** Mice at postnatal days 15–30 were anaesthetized by intraperitoneal injection of ketamine and xylazine (100 mg/kg and 10 mg/kg, respectively), perfused transcardially with cold (0–4°C) slice cutting solution ((in mM) 80 NaCl, 2.5 KCl, 1.3 NaH₂PO₄, 26 NaHCO₃, 20 D-glucose, 75 sucrose, 0.5 sodium ascorbate, 4 MgCl₂ and 0.5 CaCl₂, 315 mOsm, pH 7.4, saturated with 95% O₂/5% CO₂) and decapitated. Brains were sectioned into coronal slices of 300–400 μm in cold cutting solution with a Super Microslicer Zero1 (D.S.K.). Slices containing the NOT-DTN were incubated in a submerged chamber at 34°C for 30 min and then at room temperature (~21°C) until used for recordings. During the whole procedure, the cutting solution was bubbled with 95% O₂/5% CO₂.

Whole-cell recordings were done in ACSF (in mM: 119 NaCl, 2.5 KCl, 1.3 NaH₂PO₄, 26 NaHCO₃, 20 D-glucose, 0.5 sodium ascorbate, 4 MgCl₂, 2.5 CaCl₂, 300 mOsm, pH 7.4, saturated with 95% O₂/5% CO₂). The ACSF was warmed to ~30°C and perfused at 3 ml/min. NOT-DTN neurons were visualized with DIC infrared video-microscopy under a water immersion objective (40×, 0.8 NA) on an upright microscope (Olympus BX51WI) with an IR CCD camera (Till Photonics VX44). Whole-cell voltage-clamp recordings were performed with patch pipettes (borosilicate glass; Sutter Instruments) using a caesium-based internal solution ((in mM) 115 CsMeSO₄, 1.5 MgCl₂, 10 HEPES, 0.3 Na₃GTP, 4 MgATP, 10 Na₂-phosphocreatine, 1 EGTA, 2 QX-314-Cl, 10 BAPTA-tetracesium, 0.5% biocytin, 295 mOsm, pH 7.35). AMPA receptor-mediated EPSCs were recorded at the reversal potential for IPSCs (~−65 mV) and NMDA receptor-mediated EPSCs were recorded at +40 mV in the presence of the GABA_A receptor antagonist gabazine (5 μM, Tocris 1262) and the AMPA receptor antagonist NBQX (10 μM, Tocris 1044). To verify monosynaptic connectivity, we isolated NMDA receptor-mediated EPSCs in the presence of NBQX and high Mg²⁺ concentration (4 mM) or monosynaptic AMPA receptor-mediated EPSCs by a modified sCRACM approach⁴⁵ in the presence of tetrodotoxin (TTX; 1 μM, Tocris 1069), 4-aminopyridine (4-AP; 1.5 mM, Abcam ab120122) and tetraethylammonium (TEA; 1.5 mM, ab120275). EPSCs were acquired and filtered at 4 kHz with a Multiclamp 700B amplifier, and digitized with a Digidata 1440A at 10 kHz under the control of Clampex 10.2 (Molecular Devices). Data were analysed offline with Clampfit 10.2 (Molecular Device). To photostimulate ChR2-expressing cortico-fugal axons, we delivered blue light using a collimated LED (470 nm) and a T-Cube LED Driver (Thorlabs) through the fluorescence illuminator port and the 40× objective. Light pulses of 10 ms and 5.5 mW/mm² were given with a 20 s inter-stimulus interval. After recordings, slices were fixed by 4% paraformaldehyde for histology.

Pharmacological silencing of the NOT-DTN. After implanting the head bar, under anaesthesia (2% isoflurane), we dissected out part of the skull and removed, by aspiration, the area of the cortex and hippocampus overlaying the NOT-DTN. The identity of the NOT-DTN was assessed visually by its anatomy and stereotactic coordinates and verified electrophysiologically (see data analysis below). After the surgery, the mice were head-fixed and isoflurane was withdrawn. For at least the next 45 min, OKR performance and NOT-DTN activity were recorded. The GABA_A receptor agonist muscimol (0.2–1 mM in ACSF) was applied on top of the NOT-DTN. It took ~30 min for muscimol to silence the NOT-DTN, as assessed electrophysiologically. Pupillary dilation, as a side effect of silencing the olivary pretectal nucleus, was counteracted by topical application of 2% pilocarpine hydrochloride (agonist of muscarinic receptor, Tocris 0694) in saline to both eyes.

Histochemistry. Mice were perfused transcardially first with phosphate buffered saline (PBS, pH 7.4) and then with 4% paraformaldehyde in PBS (pH 7.4) under anaesthesia (ketamine 100 mg/kg and xylazine 10 mg/kg; intraperitoneal injection). Brains were removed from the skull, post-fixed overnight in 4% paraformaldehyde and then immersed in 30% sucrose in PBS until they sank. Brains were subsequently coronally sectioned (40–60 μm sections) with a sliding microtome (Thermo Scientific HM450). Slices were incubated in blocking buffer (PBS, 5% goat serum (Life Technologies 16210-072), 1% Triton X-100) at room temperature for 2 h and then incubated with primary antibodies in blocking buffer at 4°C overnight. The following primary antibodies were used: rabbit anti-GFP (1:1,000, Life Technologies A6455) and rabbit anti-c-Fos (1:1,000, Santa Cruz Biotechnology sc-52). The slices were washed three times with blocking buffer for 30 min each and then incubated with secondary antibodies conjugated with Alexa Fluor 488, 594 or 633 (1:800, Life Technologies A11008, A11012 or A21070, respectively) in blocking buffer for 2 h at room temperature. After being washed three times with blocking buffer for 10 min each, slices were mounted in Vectashield mounting medium containing DAPI (Vector Laboratories H1500).

For c-Fos immunostaining, 90 min after the beginning of OKR stimulation (30 min after 60-min OKR stimulation was finished), animals were perfused transcardially first with PBS and then with 4% paraformaldehyde in PBS. Brains were coronally sectioned into slices of 40 μ m.

To reveal the morphology of NOT-DTN neurons filled with biocytin, following fixation and blocking (see above), we incubated the slices with streptavidin conjugated with Alexa Fluor 647 (1:500, Life Technologies s32357) in blocking buffer overnight and then washed the slices three times.

Images were acquired on a Leica SP5 confocal microscope, a Zeiss Axio Imager A1 epifluorescence microscope or an Olympus MVX10 stereoscope, and processed using ImageJ (National Institutes of Health).

Data analysis. Analysis of eye tracking and *in vivo* electrophysiology was performed using custom-written codes in Matlab. Analysis of *in vitro* electrophysiology was done with Clampfit 10.2 (Molecular Devices).

OKR gain. Saccade-like fast eye movements were removed from the recorded eye trajectory before computing OKR amplitude (Extended Data Fig. 1c). Saccades were detected as 'spikes' in the temporal derivative of the eye position (velocity) and replaced by linear interpolation. To derive the amplitude of the OKR we used the Fourier transform of the eye position as a function of time. The eye trajectories illustrated in this study are the averages of several cycles. The gain of the OKR was expressed as $\text{OKR gain} = \text{Amp}_{\text{eye}} / \text{Amp}_{\text{drum}}$, where Amp_{eye} is the amplitude of eye movement and Amp_{drum} the amplitude of drum movement. The OKR gain derived in the space domain is similar to that derived in the velocity domain (Extended Data Fig. 1f). In this study, we computed the gain in the space domain because deriving eye velocity from eye position introduces noise. Therefore, the OKR gain is 1 if the eye perfectly tracks the trajectory of the virtual drum and 0 if it does not track.

Cortical contribution to OKR gain. The cortical contribution to the OKR gain is expressed as the percentage reduction in OKR gain caused by cortical silencing and calculated as $\Delta V (\%) = (V_{\text{control}} - V_{\text{silencing}}) / V_{\text{control}}$, where V_{control} and $V_{\text{silencing}}$ are the values of the OKR gain measured under control conditions or during optogenetic cortical silencing, respectively.

OKR potentiation following vestibular lesion. OKR potentiation is calculated as $V_{\text{post VL}} / V_{\text{pre VL}}$, where $V_{\text{pre VL}}$ and $V_{\text{post VL}}$ are the values of the OKR gain measured before and after vestibular lesion, respectively.

Cortical contribution to OKR potentiation. The cortical contribution to OKR potentiation is expressed as $\text{PI} = (\Delta V_{\text{post VL}} - \Delta V_{\text{pre VL}}) / (\Delta V_{\text{max}} - \Delta V_{\text{pre VL}})$, where $\Delta V_{\text{pre VL}}$ and $\Delta V_{\text{post VL}}$ are the cortical contribution to the OKR gain before and after vestibular lesioning, respectively, and ΔV_{max} is the maximum possible cortical contribution to the OKR gain assuming that the entire amount of OKR potentiation depends on visual cortex. $\Delta V_{\text{max}} = (V_{\text{post VL, control}} - V_{\text{pre VL, silencing}}) / V_{\text{post VL, control}}$. Hence PI is 1 if the entire amount of OKR potentiation depends on visual cortex and is 0 if the cortical contribution to OKR gain before vestibular lesion is the same as the cortical contribution to OKR gain after vestibular lesion ($\Delta V_{\text{pre VL}} = \Delta V_{\text{post VL}}$) (Extended Data Fig. 3c, d).

Cortical contribution to NOT-DTN activity. The cortical contribution to NOT-DTN activity is expressed as the cortical contribution to OKR gain but V_{control} and $V_{\text{silencing}}$ are the firing rates of NOT-DTN neurons under control conditions or during optogenetic cortical silencing, respectively.

Unit isolation. Single units were isolated using spike-sorting Matlab codes, as described previously⁴³. The raw extracellular signal was band-pass filtered between 0.5 and 10 kHz. Spiking events were detected with a threshold at 3.5 or 4 times the standard deviation of the filtered signal. Spike waveforms of four adjacent electrode sites were clustered using a *k*-means algorithm. After initial automated clustering, clusters were manually merged or split with a graphical user interface in Matlab. Unit isolation quality was assessed by considering refractory period violations and Fisher linear discriminant analysis. All units were assigned a depth according to the electrode sites at which their amplitudes were largest. Multi-unit spiking activity was defined as all spiking events exceeding the detection threshold after the removal of electrical noise or movement artefacts by the sorting algorithm. Individual spiking events were also assigned to one of the 16 recording sites according to where they showed the largest amplitude. For both single-unit activity and multi-unit activity, the visual response was computed as the mean firing rate during visual stimulation without baseline subtraction.

Units recorded from visual cortex were assigned as regular-spiking neurons or fast-spiking putative inhibitory neurons based on the trough-to-peak times of spike waveforms⁴³. A threshold of 0.4 ms was used to distinguish fast-spiking from regular-spiking units.

Direction selectivity index. The boundary of the NOT-DTN was determined by the appearance of a temporonasal directional bias in the multi-unit response to the visual stimulus.

The preferred direction of an isolated NOT-DTN unit was determined by summing response vectors of 12 evenly spaced directions. The direction selectivity index (DSI) was calculated along the sampled orientation axis closest to the preferred direction according to the formula $\text{DSI} = (R_{\text{pref}} - R_{\text{null}}) / (R_{\text{pref}} + R_{\text{null}})$, where R_{pref} is the response at the preferred direction and R_{null} is the response at the opposite direction.

The DSI of the response evoked by oscillatory drum movement was calculated as $\text{DSI} = (R_{\text{TN}} - R_{\text{NT}}) / (R_{\text{TN}} + R_{\text{NT}})$, where R_{TN} is the response during the temporonasal phase of drum movement and R_{NT} is the response during the nasotemporal phase. **Onset latency and jitter.** The onset latency of optogenetically evoked activity of NOT-DTN neurons was determined as the time lag between the beginning of the LED illumination and the time point at which the firing rate reached three times the standard deviation of spontaneous activity. Similarly, the onset latency of optogenetically evoked EPSCs in NOT-DTN neurons was determined as the time lag between the beginning of the LED illumination and the time point at which the EPSC amplitude reached three times the standard deviation of baseline noise. Trial-by-trial jitter of optogenetically evoked EPSCs was calculated as the standard deviation of the onset latency.

Overlap coefficient. Analysis of c-Fos immunohistochemistry was performed with ImageJ (National Institutes of Health). c-Fos-positive cells were identified as continuous pixels after thresholding and counted automatically. To quantify the extent of overlap between arborization of GFP-expressing RGC axons and c-Fos expression in the NOT-DTN, their boundaries were manually drawn and the overlap coefficient r was calculated as

$$r = \frac{\sum (S1_i \times S2_i)}{\sqrt{\sum (S1_i)^2 \times \sum (S2_i)^2}}$$

where $S1_i$ is 1 if pixel i is within the domain of RGC axons, otherwise 0; and $S2_i$ is 1 if pixel i is within the domain of c-Fos immunohistochemistry, otherwise 0 (Extended Data Fig. 5c).

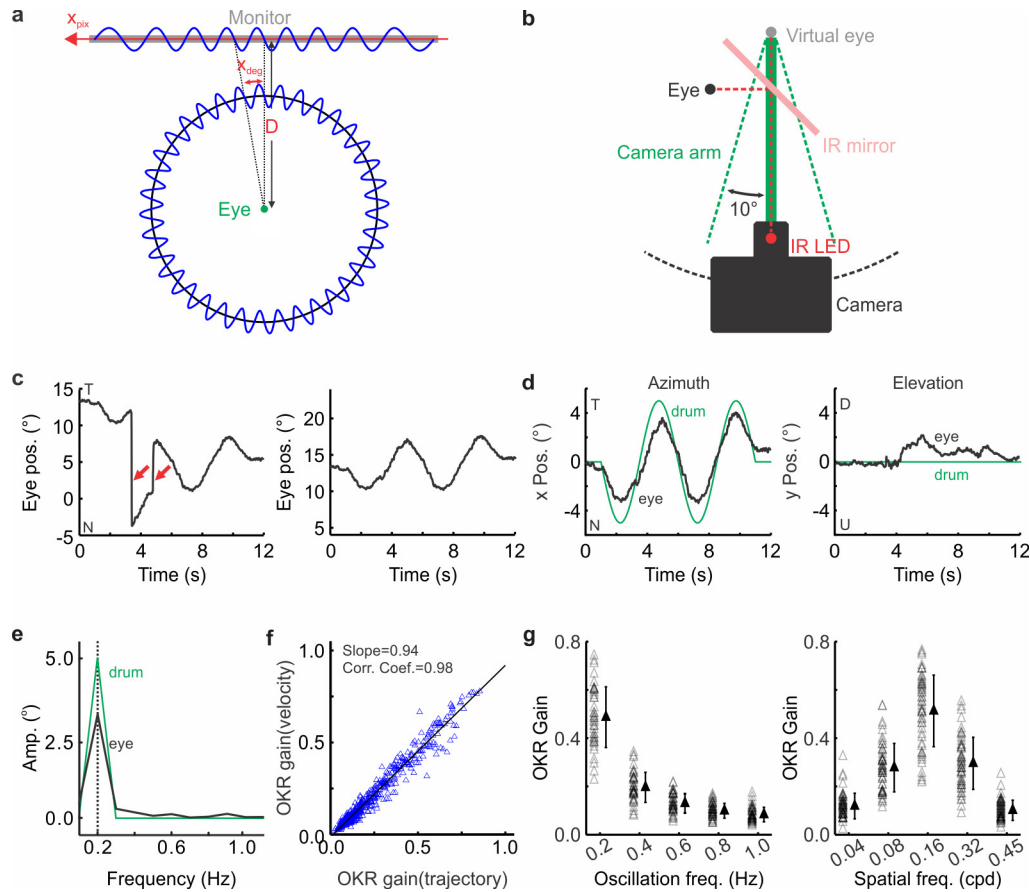
Averaged normalized transfer function. For each animal, NOT-DTN multiunit activity was normalized to the average firing rate evoked by optimal spatial frequency. Data points of transfer functions from all animals were pooled, binned and averaged.

Vector analysis of the effect of cortical silencing on the transfer function. The vectors (arrows in Extended Data Fig. 9g–i) start at the centre of mass of data points obtained at a given spatial frequency under control conditions (grey) and end at the centre of mass of data points obtained at the same spatial frequency during cortical silencing trials (blue). The x-axis value of the centre of mass is the NOT-DTN multiunit firing rate averaged over trials obtained at a given spatial frequency, normalized by the average firing rate evoked by the best spatial frequency. The y-axis value of the centre of mass is the average OKR gain obtained during the same trials.

Inclusion/exclusion criteria. All samples or animals were included in the analysis except for the following exclusions: (1) in the analysis of OKR gain, trials in which video-oculography failed as a result of eye blinking or tears were excluded from analysis; (2) in Fig. 1g, h, one mouse was excluded from the analysis because its value of OKR potentiation was less than the threshold of 0.1; (3) in Fig. 3, two mice were excluded from the analysis because they were sick and lost a lot in body weight during experiments; (4) in Figs. 4, 5, one mouse was excluded because the identification of NOT-DTN failed; and (5) in statistics of the activities of superior colliculus and vLGN, recordings which were identified *post hoc* as missing the target structures were excluded from the analysis. These criteria were pre-established.

Statistical analysis. Statistical analyses were done using statistics toolbox in Matlab. All data are presented as mean \pm s.e.m. unless otherwise noted. Statistical significance was assessed using paired or unpaired *t*-tests and further confirmed with nonparametric Wilcoxon signed rank test or Wilcoxon rank sum test unless otherwise noted. Estimated sample sizes were retrospectively determined to achieve 80% power to detect expected effect sizes using Matlab. We did not intentionally select particular mice for treatment group or control group. No blinding was used. Owing to the limited sample size, the assumption of normal distribution was not tested. Nonparametric tests were used to confirm statistical significances reported by paired or unpaired *t*-tests. Thus, the conclusions of statistical tests were validated regardless of whether the data were normally distributed. The variance was not compared between groups. In *t*-tests, we assumed that samples were from distributions of unknown and unequal variances. The experiments were not randomized.

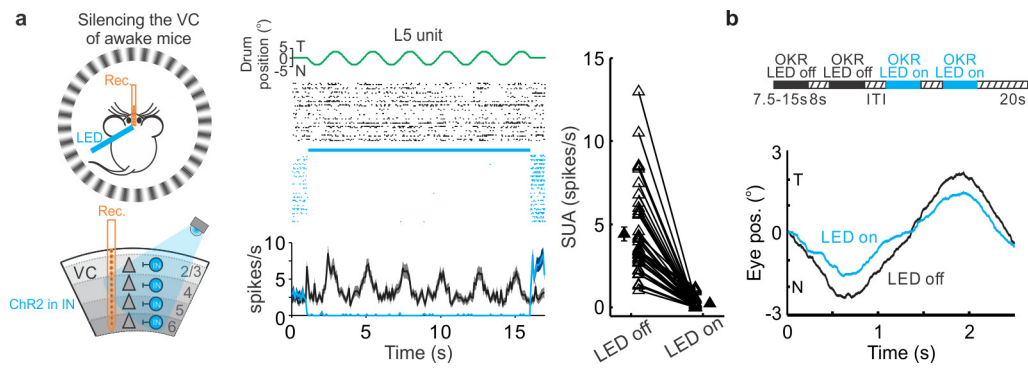
31. Zhao, S. *et al.* Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat. Methods* **8**, 745–752 (2011).
32. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
33. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
34. Dhande, O. S. *et al.* Genetic dissection of retinal inputs to brainstem nuclei controlling image stabilization. *J. Neurosci.* **33**, 17797–17813 (2013).
35. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100**, 13940–13945 (2003).
36. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
37. Azim, E., Jiang, J., Alstermark, B. & Jessell, T. M. Skilled reaching relies on a V2a propriospinal internal copy circuit. *Nature* **508**, 357–363 (2014).
38. Soudais, C., Laplace-Builhe, C., Kissa, K. & Kremer, E. J. Preferential transduction of neurons by canine adenovirus vectors and their efficient retrograde transport in vivo. *FASEB J.* **15**, 2283–2285 (2001).
39. Prusky, G. T., Alam, N. M., Beekman, S. & Douglas, R. M. Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. *Invest. Ophthalmol. Vis. Sci.* **45**, 4611–4616 (2004).
40. Collewijn, H. Optokinetic eye movements in the rabbit: input-output relations. *Vision Res.* **9**, 117–132 (1969).
41. Stahl, J. S., van Alphen, A. M. & De Zeeuw, C. I. A comparison of video and magnetic search coil recordings of mouse eye movements. *J. Neurosci. Methods* **99**, 101–110 (2000).
42. Stahl, J. S. Calcium channelopathy mutants and their role in ocular motor research. *Ann. NY Acad. Sci.* **956**, 64–74 (2002).
43. Olsen, S. R., Bortone, D. S., Adesnik, H. & Scanziani, M. Gain control by layer six in cortical circuits of vision. *Nature* **483**, 47–52 (2012).
44. Liu, B. H. *et al.* Intervening inhibition underlies simple-cell receptive field structure in visual cortex. *Nat. Neurosci.* **13**, 89–96 (2010).
45. Petreanu, L., Mao, T., Sternson, S. M. & Svoboda, K. The subcellular organization of neocortical excitatory connections. *Nature* **457**, 1142–1145 (2009).



Extended Data Figure 1 | Quantification of mouse OKR.

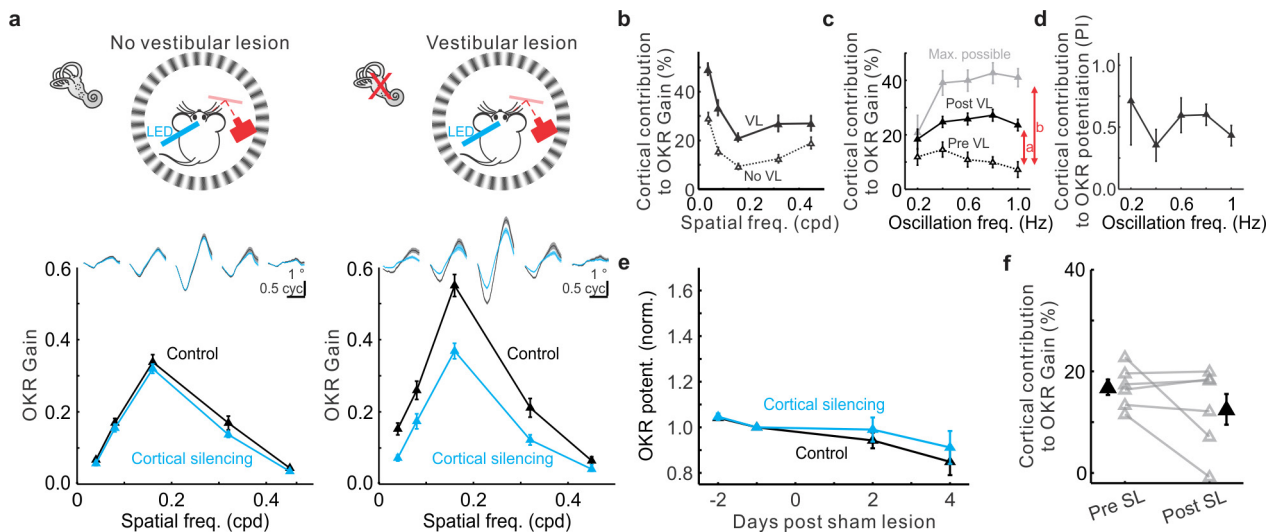
a, Transformation of sinusoidal gratings from cylindrical coordinates of the virtual drum to Cartesian coordinates of the monitor. x_{pix} is the horizontal pixel position in Cartesian coordinates. D is the distance from the centre of monitors to the eye. x_{deg} is the azimuth angle of pixels in cylindrical coordinates. Note that the spatial period of the grating on the monitor is not uniform. See Methods for details. **b**, Schematic of calibration of the measurement of eye position. The camera is moved along a circumference centred on the image of the eye by $\pm 10^\circ$. **c–e**, Example traces of OKR eye trajectory and corresponding fast Fourier transform (FFT) spectra. **c**, Left, raw trace of one individual eye trajectory with both slow OKR component and fast saccade-like component (red arrows; T, temporal; N, nasal). Right, isolated OKR component after removal of the saccade-like component. Spatial frequency, 0.08 cpd;

oscillation frequency, 0.2 Hz. **d**, Eye trajectories in horizontal azimuth (left) and vertical elevation (right) overlaid with corresponding drum trajectories (the same example as in **c**). Note that OKR eye movement is mainly restricted to the axis of the drum movement. D, down; U, up. **e**, Fourier transform spectra of eye trajectory and drum trajectory in **d** (left). The amplitude of the OKR trajectory peaks at the principal frequency (dotted line). **f**, OKR gain derived from OKR velocity versus OKR eye trajectory. Each point is one trial. Solid line, linear regression. **g**, Population summary of OKR gain evoked by five oscillation frequencies (left, spatial frequency 0.08 cpd) and five spatial frequencies (right, oscillation frequency 0.4 Hz). Each point is one mouse ($n = 39$ for oscillation frequency and 49 for spatial frequency). Data shown as mean \pm s.d.



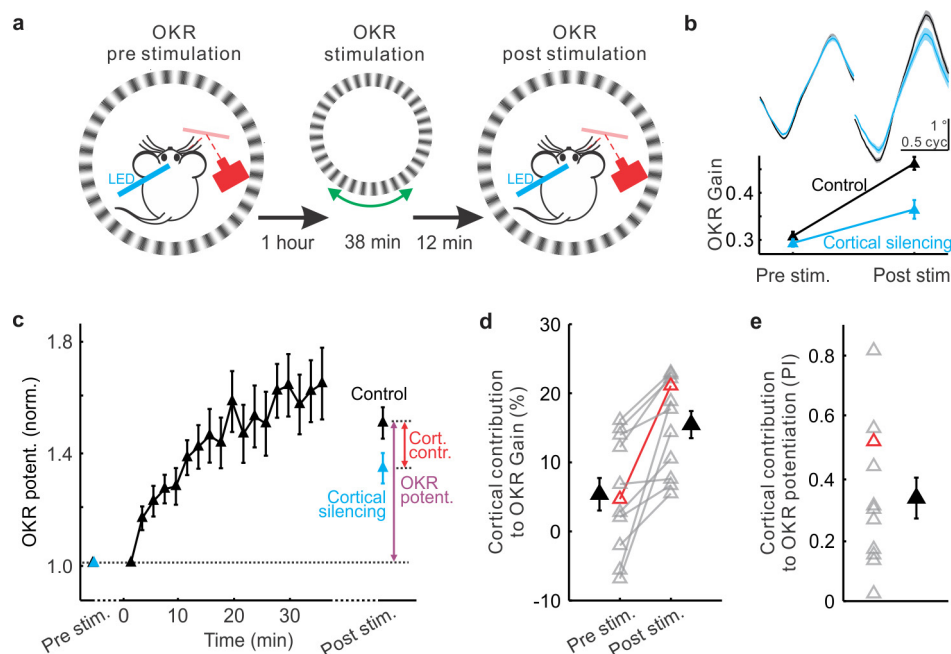
Extended Data Figure 2 | Optogenetic silencing of visual cortex. **a**, Left, schematic of experimental setup. IN, inhibitory neurons; VC, visual cortex. Middle, raster plot and PSTH of a single unit. Black, control condition; blue, cortical silencing. Blue bar, duration of blue light illumination (15 s). Control and photostimulation trials were interleaved

(see **b**), but are separated here for clarity. Right, summary of firing rate of regular spiking units ($n = 40$). Data shown as mean \pm s.e.m. **b**, Top, block design to examine the impact of cortical silencing on OKR performance. LED off, control trials; LED on, cortical silencing trials. Bottom, cycle averages of one individual OKR eye trajectory. T, temporal; N, nasal.



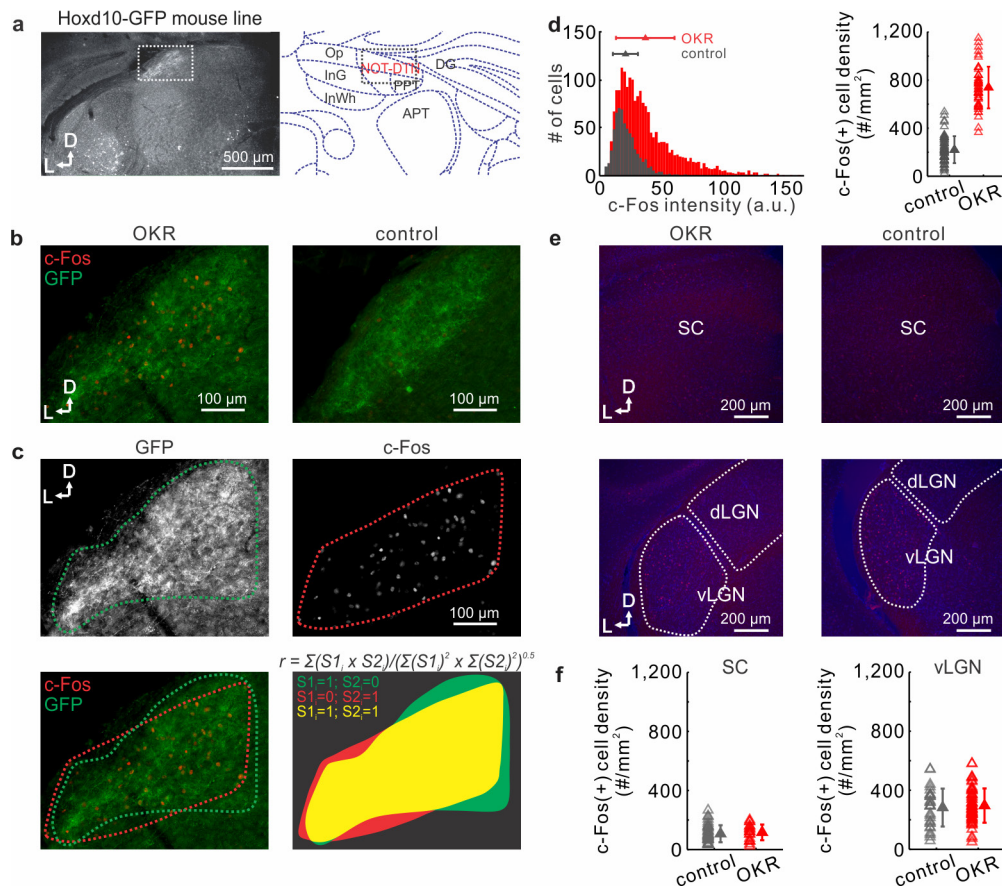
Extended Data Figure 3 | Visual cortex contributes to OKR potentiation across spatial frequencies. **a**, Data from example mice. OKR performance of a naive animal (no vestibular lesion, left) and an animal with vestibular lesion (right). Top, schematic experimental setup. Bottom, cycle averages of all eye trajectories evoked by five spatial frequencies (oscillation frequency 0.4 Hz), and the corresponding OKR gains. Thickness of traces shows s.e.m. Data shown as mean \pm s.e.m. **b**, Population average of cortical contribution to OKR gain at five spatial frequencies for animals with vestibular lesion (VL, solid line, $n=17$ mice) and naive animals (no VL, dotted line, $n=51$ mice). Data shown as mean \pm s.e.m. **c**, Population average of cortical contribution to OKR gain at five oscillation frequencies before vestibular lesion (Pre VL, dotted line) and after vestibular lesion (Post VL, solid black line). The grey line is the

maximal possible cortical contribution to OKR gain after vestibular lesion assuming the entire OKR potentiation depends on visual cortex (Max. possible) ($n=13$ animals). Data shown as mean \pm s.e.m. **d**, Population average of cortical contribution to OKR potentiation (potentiation index, PI) measured as the ratio between *a* and *b* (illustrated in **c**) at each oscillation frequency. Data shown as mean \pm s.e.m. **e**, Population averages of pseudo-OKR potentiation following sham lesions. Black data points: no cortical silencing, normalized by OKR gain before sham lesions without cortical silencing. Blue data points: cortical silencing, normalized by OKR gain before sham lesions during cortical silencing ($n=6$ mice). Data shown as mean \pm s.e.m. **f**, Population summary of cortical contribution to OKR gain before (Pre SL) and after sham lesions (Post SL). Data shown as mean \pm s.e.m.



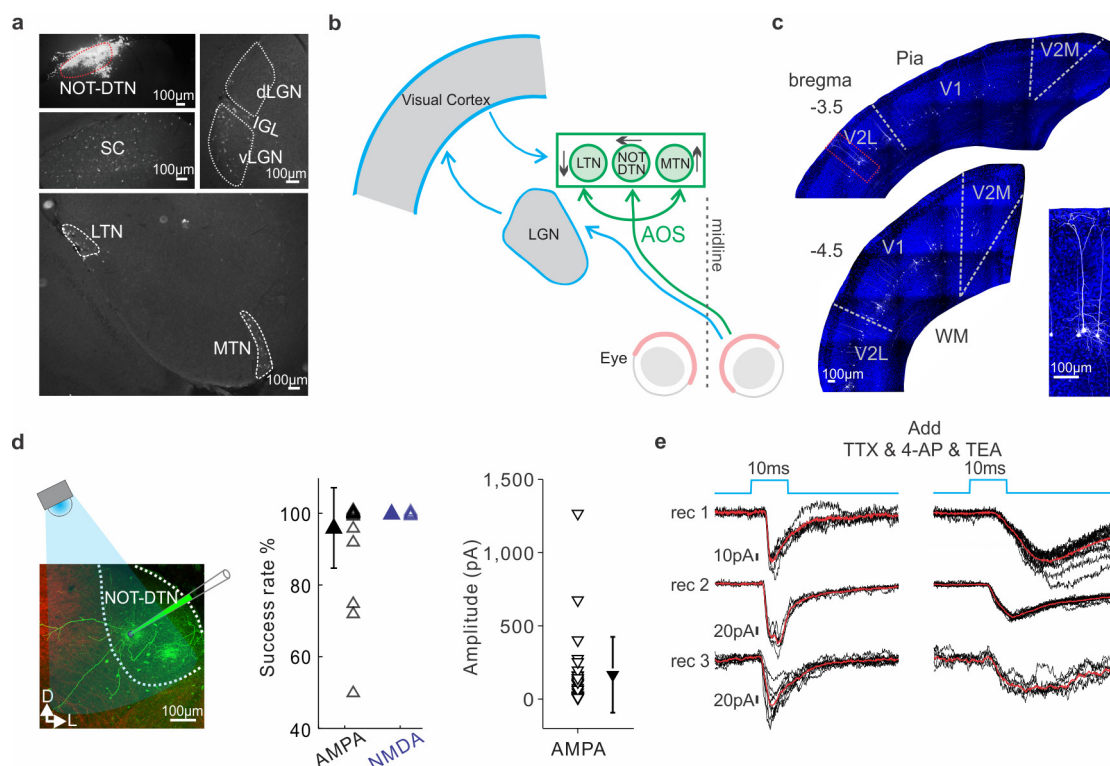
Extended Data Figure 4 | Visual cortex contributes to OKR potentiation induced by continuous OKR stimulation. **a**, Schematic of experimental design. OKR gain before stimulation was measured twice, 1 day before and 1 h before continuous OKR stimulation (see Methods for details). **b**, Data from example mouse. Cycle averages of all eye trajectories and corresponding OKR gain before (Pre stim.) and after (Post stim.) continuous OKR stimulation ($n = 576$ cycles, spatial frequency 0.1 cpd, oscillation frequency 0.4 Hz). The thickness of the trace shows s.e.m. Note that following OKR stimulation cortical silencing leads to a larger reduction in OKR gain. Data shown as mean \pm s.e.m. **c**, Population averaged time course of OKR potentiation induced by continuous OKR

stimulation. Black, no cortical silencing (control), normalized by OKR gain before stimulation (Pre stim.) without cortical silencing. Blue, cortical silencing, normalized by OKR gain before stimulation during cortical silencing ($n = 11$ mice). Red arrow: the cortical contribution to OKR potentiation; magenta arrow: OKR potentiation. Data shown as mean \pm s.e.m. **d**, Population summary of cortical contribution to OKR gain before (Pre stim.) and after stimulation (Post stim.) ($n = 11$ mice). Red data points: the animal in **b**. Data shown as mean \pm s.e.m. **e**, Population summary of cortical contribution to OKR potentiation (potentiation index, PI) ($n = 11$ mice). Red data point: the animal in **b**. Data shown as mean \pm s.e.m.



Extended Data Figure 5 | Identification of the NOT-DTN based on retinal input and c-Fos expression. **a**, Left, coronal section of NOT-DTN of Hoxd10-GFP mouse. The distribution of GFP-expressing RGC axons delineates the NOT-DTN (dotted box). Right, delineation of NOT-DTN and surrounding nuclei (modified from Paxinos, G. & Franklin, K. *The Mouse Brain in Stereotaxic Coordinates* (Elsevier, 2007)) for the corresponding coronal plane. D, dorsal; L, lateral. **b**, c-Fos immunostaining of coronal slices containing NOT-DTN of Hoxd10-GFP mice. Left, section from an animal that underwent OKR stimulation. Note that the distribution of GFP RGC axons overlaps with that of c-Fos-positive cells. Right, section from an animal that did not undergo OKR stimulation (control). **c**, Quantification of the extent of overlap between GFP RGC axons and c-Fos-positive cells in **b** (left). Top left,

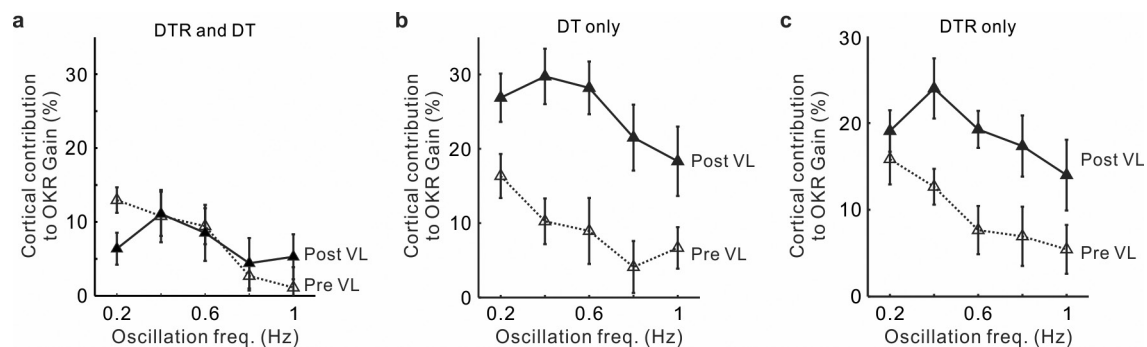
boundary of the domain of RGC axons. Top right, boundary of the domain of c-Fos-positive cells. Bottom left, overlay of those two boundaries. Bottom right, calculation of overlap coefficient r of those two domains (see Methods). **d**, Left, histogram of fluorescence intensity of c-Fos-positive cells. Data shown as mean \pm s.d. $P < 10^{-20}$. Right, summary of density of c-Fos-positive cells in NOT-DTN. Each data point represents one slice. Data shown as mean \pm s.d. $P < 10^{-20}$. $n = 50$ slices from 4 mice of OKR group and 59 slices from 4 mice of control group. **e**, c-Fos immunostaining of coronal slices containing superior colliculus (SC, top) or vLGN (bottom). Blue, DAPI; red, c-Fos. **f**, Summary of density of c-Fos-positive cells in superior colliculus (left) and vLGN (right). Each data point represents one slice. Data shown as mean \pm s.d. $n = 4$ mice for both OKR group and control group.



Extended Data Figure 6 | Structures projecting to NOT-DTN and monosynaptic transmission between visual cortex and NOT-DTN.

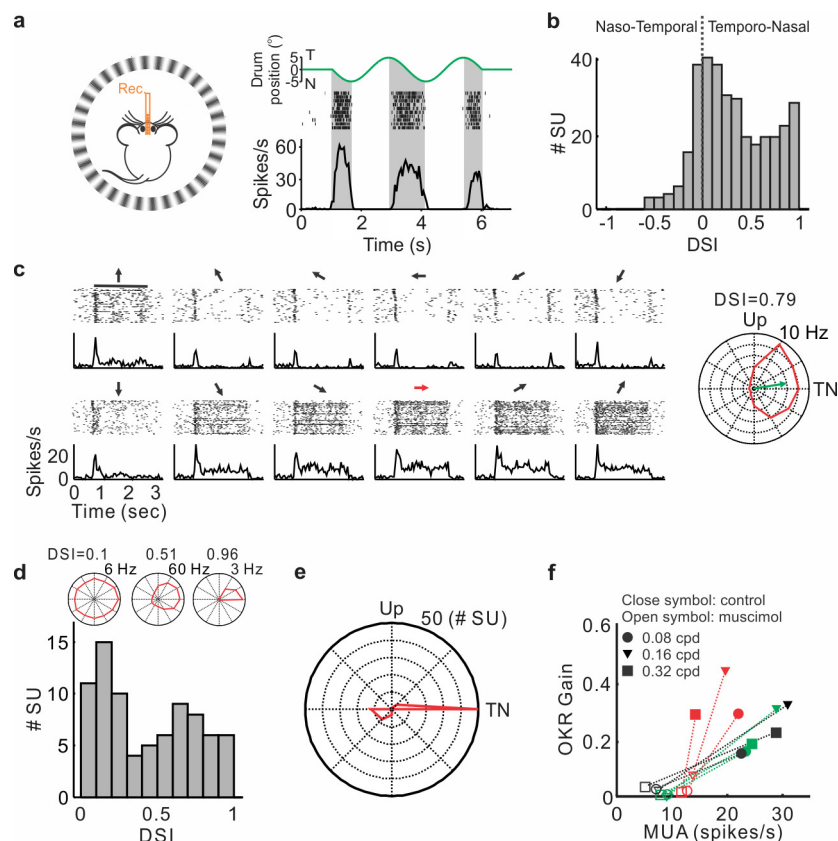
a, Subcortical structures labelled by retro-beads injected into the NOT-DTN. Top left, injection site. SC, superior colliculus; dLGN, dorsal lateral geniculate nucleus; IGL, intergeniculate leaflet; vLGN, ventral lateral geniculate nucleus; LTN, lateral terminal nucleus; MTN, medial terminal nucleus. **b**, Schematic drawing of the two pathways relaying visual information to the AOS. Thalamo-cortical-NOT-DTN pathway is outlined in blue, retinal pathway outlined in green. **c**, Spatial distribution of NOT-DTN-projecting neurons in visual cortex (visual cortex injected with Flex-tdTomato and NOT-DTN with Cav2-Cre) for two coronal sections. Boundaries between primary and secondary areas are drawn according to Paxinos, G. & Franklin, K. *The Mouse Brain in Stereotaxic Coordinates* (Elsevier, 2007). Inset on the right, higher magnification of the region

shown in the red box. Blue, DAPI; white, tdTomato. **d**, Left, schematic of the setup for *in vitro* whole-cell voltage-clamp recording from NOT-DTN neurons in acute slices. Green, patched NOT-DTN neurons; red, axons from visual cortex. D, dorsal; L, lateral. Middle, summary of success rate of EPSCs evoked by optogenetic stimulation of cortico-fugal axons. Right, peak amplitude of AMPA receptor mediated EPSCs. Data shown as mean \pm s.d. Each data point represents one NOT-DTN recording. **e**, Left, AMPA receptor-mediated EPSCs evoked by optogenetic stimulation of cortico-fugal axons for three NOT-DTN neurons voltage-clamped at -65 mV. Right, AMPA receptor-mediated EPSCs of the same cells after blocking multi-synaptic components with TTX (sCRACM). Black, individual trials; red, average; blue, time course of blue light illumination.



Extended Data Figure 7 | Cortical contribution to OKR gain at different oscillation frequencies in animals with spared or ablated cortical projection to the NOT-DTN. **a**, Population averages of cortical contribution to OKR gain at five different oscillation frequencies before (dotted line, Pre VL) and after (solid line, Post VL) vestibular lesion for mice in which the cortico-fugal projection was ablated ($n = 18$ animals).

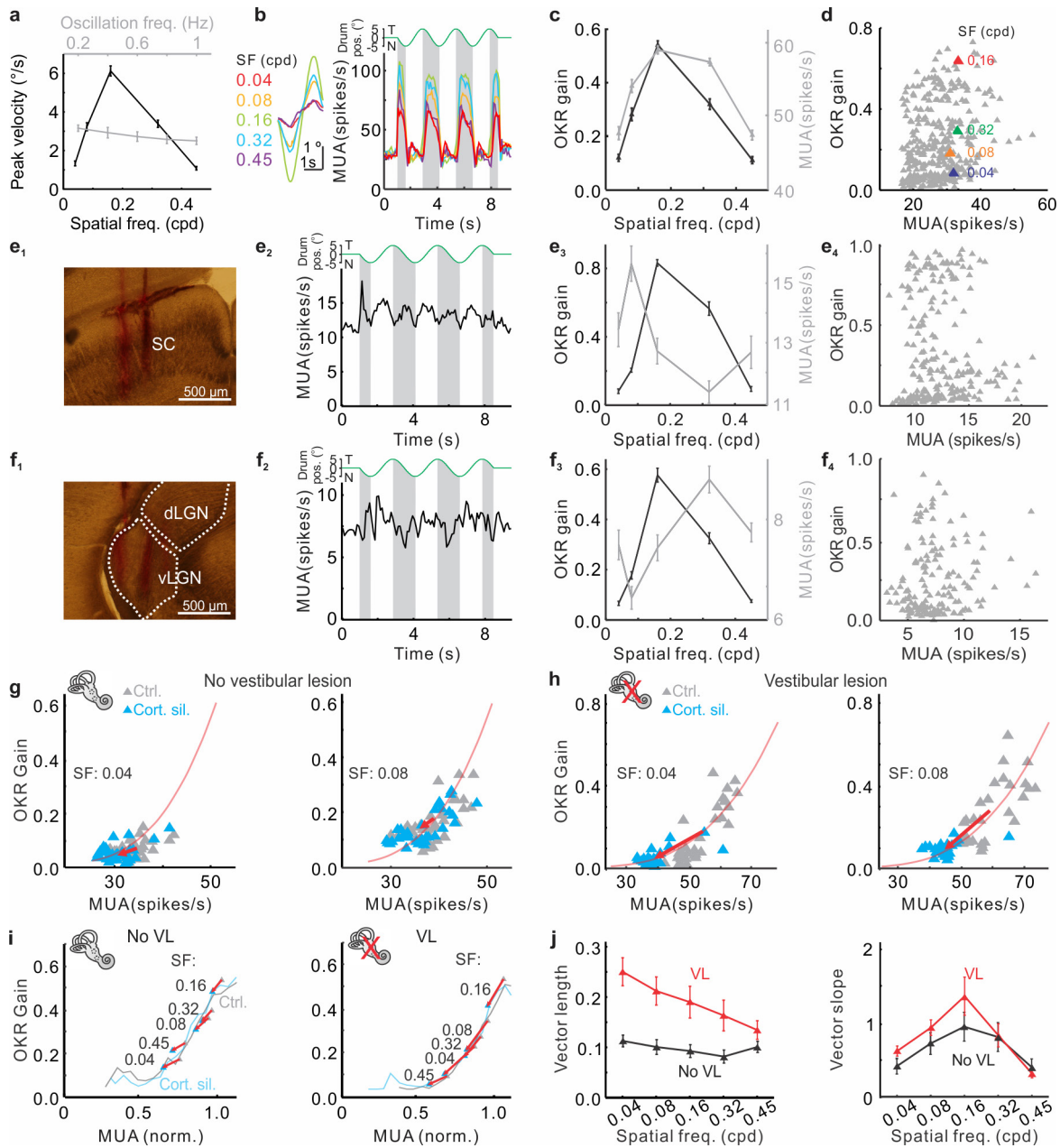
Data shown as mean \pm s.e.m. **b, c**, Population averages of cortical contribution to OKR gain at five different oscillation frequencies for mice in which the infection with diphtheria toxin receptor (DTR) (**b**) or injection of diphtheria toxin (DT) (**c**) was omitted ($n = 8$ and 6 animals, respectively). Data shown as mean \pm s.e.m.



Extended Data Figure 8 | Tuning properties of NOT-DTN neurons.

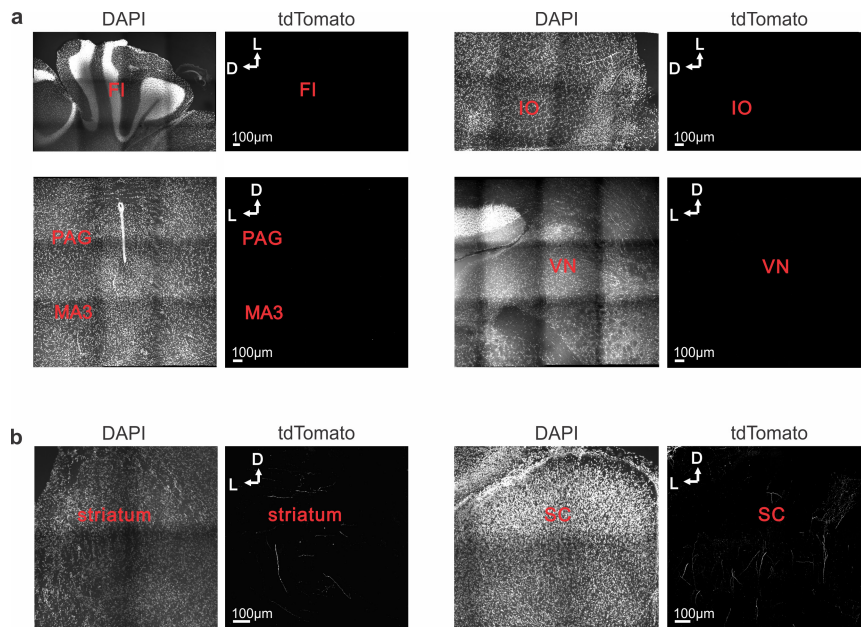
a, Left, schematic of experimental setup; mouse under anaesthesia. Right, raster plot and PSTH of an example single unit. Shaded regions indicate the temporonasal phase of drum trajectory. **b**, Histogram of direction selectivity index (DSI) of single units in NOT-DTN stimulated by oscillatory drum movement. **c**, Example single unit. Left, raster plot and PSTH of responses evoked by moving gratings of 12 equally spaced directions (indicated by arrows, red arrow for temporonasal direction). Bar, duration of stimulation. Right, polar plot of the same unit. Green

arrow, preferred direction. **d**, Top, example polar plots of weak DSI, medium DSI and strong DSI units. Bottom, histogram of DSI of NOT-DTN units stimulated by grating movement of 12 directions. **e**, Summary of preferred direction for NOT-DTN units with DSI greater than 0.1. Note the dominant preference for temporonasal direction. **f**, OKR gain and NOT-DTN multi-unit activity recorded before (closed) and after (open) silencing NOT-DTN with muscimol. Each colour represents one animal. Note that strong suppression of NOT-DTN activity leads to the abolishment of the OKR. Mice were awake during recording.



Extended Data Figure 9 | Cortical silencing induces a larger shift along the transfer function after vestibular lesion. **a**, Example spatial frequency tuning and oscillation frequency tuning curves of OKR peak velocity. Note that while the OKR peak velocity is modulated by the spatial frequency of the drum stimulus (black), it is constant across oscillation frequencies (grey). Data shown as mean \pm s.e.m. **b**, Data from example mouse. Left, cycle averages of all eye trajectories triggered by five different spatial frequencies. Right, the corresponding PSTH of NOT-DTN multi-unit activity. Note the correlation between the amplitude of eye trajectory and the amplitude of activity. Shades indicate the temporonasal phase of drum trajectory. **c**, Spatial frequency tuning curves of OKR gain (black) and NOT-DTN activity (grey) from **b**. Data shown as mean \pm s.e.m. **d**, Pseudo-transfer function from the animal shown in Fig. 5b using the firing rate during the nasotemporal instead of the temporonasal phase. Each data point represents one trial. Coloured triangles represent the same trials as illustrated in Fig. 5b (middle). Note the lack of correlation between OKR gain and the nasotemporal phase of multi-unit activity (MUA) recorded in NOT-DTN. **e**, Example spiking activity in superior colliculus (SC) during OKR stimulation. **e₁**, Image of coronal slice containing superior colliculus. Red, electrode track labelled with DiI. **e₂**, PSTH of superior colliculus MUA. **e₃**, Spatial frequency tuning curves of OKR gain (black)

and superior colliculus activity (grey). Data shown as mean \pm s.e.m. **e₄**, Pseudo-transfer function using superior colliculus activity. Note the lack of correlation between OKR gain and MUA recorded in superior colliculus. **f**, As in **e**, except for spiking activity in ventral lateral geniculate nucleus (vLGN) during OKR stimulation. Note the lack of correlation between OKR gain and MUA recorded in vLGN. Data shown as mean \pm s.e.m. **g**, Data from example mouse. Recording from NOT-DTN. Shift along the transfer function upon cortical silencing for data points obtained at two different spatial frequencies (SF; left, 0.04 cpd; right, 0.08 cpd) in a naive animal (no vestibular lesion). The vector (arrow) connects the centres of mass of control (grey) and cortical silencing trials (blue) obtained at the same spatial frequency. Red line, transfer function computed with data obtained at all tested spatial frequencies under control conditions (that is, without cortical silencing). **h**, As in **g**, except for an animal with vestibular lesion. Note longer vectors as compared to **g**. **i**, Population summary of vectors for five different spatial frequencies computed on averaged normalized transfer functions in naive animals (no VL; left; $n = 17$) and animals with vestibular lesion (VL; right; $n = 17$). **j**, Population averages of vector lengths (left) and slopes (right) for naive animals (no VL; black; $n = 17$) and animals with vestibular lesion (VL; red; $n = 17$). Data shown as mean \pm s.e.m.



Extended Data Figure 10 | Presence and absence of collaterals from NOT-DTN-projecting cortical neurons in selected brain areas.

a, Absence of collaterals from NOT-DTN-projecting cortical neurons in flocculus (FL); inferior olive (IO); periaqueductal grey (PAG); medial accessory oculomotor nucleus (MA3); and vestibular nuclei (VN).

For each coronal section the left panel is the DAPI fluorescence signal (blue channel) and the right panel is the tdTomato fluorescence signal (red channel). **b**, Presence of collaterals from NOT-DTN-projecting cortical neurons in the striatum and superior colliculus (SC).

Allogeneic transplantation of iPSC cell-derived cardiomyocytes regenerates primate hearts

Yuji Shiba^{1,2*}, Toshihito Gomibuchi^{3*}, Tatsuhiro Seto³, Yuko Wada³, Hajime Ichimura³, Yuki Tanaka³, Tatsuki Ogasawara³, Kenji Okada³, Naoko Shiba⁴, Kengo Sakamoto⁵, Daisuke Ido⁵, Takashi Shiina⁶, Masamichi Ohkura^{7,8}, Junichi Nakai^{7,8}, Narumi Uno⁹, Yasuhiro Kazuki⁹, Mitsuo Oshimura⁹, Itsunari Minami¹⁰ & Uichi Ikeda²

Induced pluripotent stem cells (iPSCs) constitute a potential source of autologous patient-specific cardiomyocytes for cardiac repair, providing a major benefit over other sources of cells in terms of immune rejection. However, autologous transplantation has substantial challenges related to manufacturing and regulation. Although major histocompatibility complex (MHC)-matched allogeneic transplantation is a promising alternative strategy¹, few immunological studies have been carried out with iPSCs. Here we describe an allogeneic transplantation model established using the cynomolgus monkey (*Macaca fascicularis*), the MHC structure of which is identical to that of humans. Fibroblast-derived iPSCs were generated from a MHC haplotype (HT4) homozygous animal and subsequently differentiated into cardiomyocytes (iPSC-CMs). Five HT4 heterozygous monkeys were subjected to myocardial infarction followed by direct intra-myocardial injection of iPSC-CMs. The grafted cardiomyocytes survived for 12 weeks with no evidence of immune rejection in monkeys treated with clinically relevant doses of methylprednisolone and tacrolimus, and showed electrical coupling with host cardiomyocytes as assessed by use of the fluorescent calcium indicator G-CaMP7.09. Additionally, transplantation of the iPSC-CMs improved cardiac contractile function at 4 and 12 weeks after transplantation; however, the incidence of ventricular tachycardia was transiently, but significantly, increased when compared to vehicle-treated controls. Collectively, our data demonstrate that allogeneic iPSC-CM transplantation is sufficient to regenerate the infarcted non-human primate heart; however, further research to control post-transplant arrhythmias is necessary.

Human iPSCs, like human embryonic stem (ES) cells, are a promising cell source for cardiac repair because of their unlimited self-renewal and ability to differentiate into cardiomyocytes^{2–4}. Theoretically, human iPSCs could be used for autologous transplantation; however, it is not clear whether this strategy will be feasible in a clinical setting because it is time-consuming, laborious, and costly. This is particularly true for heart regeneration, which requires a large number of cells². Allogeneic transplantation of iPSC-CMs could solve these practical issues. A potential disadvantage of allogeneic transplantation is that it can induce an immune response, which might cause graft rejection. The MHC plays an essential role in the post-transplant immune response⁵ and graft-versus-host disease⁶; therefore, the use of MHC-matched transplants is a potential approach to avoid rejection. In the present study, we investigated whether MHC-matched allogeneic iPSC-CMs can survive in the long term following transplantation without forming tumours in a non-human primate myocardial infarction model. Additionally, we assessed the mechanical and electrical consequences of transplantation in this clinically relevant model.

We screened the MHC RNA sequences of Filipino cynomolgus monkeys and identified an animal with strictly homozygous MHC-class I (ref. 7) and MHC-class II (ref. 8) regions on both chromosomes (named HT4, Extended Data Fig. 1a, b). We designated this HT4 homozygous animal as an iPSC donor and isolated skin fibroblasts from the animal. iPSCs were established by transfection with plasmid vectors encoding *OCT4* (also known as *POU5F1*), *SOX2*, *KLF4* and *L-MYC* (*MYCL*) and subsequently formed typical ES-cell-like colonies (Extended Data Fig. 2a), expressed pluripotent markers (Extended Data Fig. 2b–f) and displayed the ability to form teratomas (Extended Data Fig. 2g–i). We previously used a human ES cell line expressing the fluorescent calcium indicator GCaMP3 to show that grafted cardiomyocytes could couple with host cardiomyocytes in an injured guinea-pig heart^{9,10}; however, we were unable to detect sufficient fluorescent signals from GCaMP3-expressing iPSC-CMs in a monkey heart in our system (data not shown), suggesting a need for an indicator with enhanced fluorescence. As such, we developed a fluorescent calcium indicator, G-CaMP7.09 (Extended Data Fig. 3a), which was transfected into undifferentiated cynomolgus monkey iPSCs. After expansion, the majority of G-CaMP7.09-expressing cells showed fluorescence, and 14 out of 15 metaphases displayed a normal karyotype (42, XY; Extended Data Fig. 2j). Next, we generated iPSC-CMs using our previously reported protocol^{9,11} as modified by another group¹². Because incubation of iPSC derivatives in glucose-free medium¹³ for 3 days significantly increased the fraction of cardiomyocytes ($P < 0.01$) in a preliminary experiment (Extended Data Fig. 4a), we added this selection step following cardiac differentiation in the present study (Extended Data Fig. 5a). We prepared 4×10^8 cardiomyocytes for a recipient animal (cardiac troponin T (cTnT)-positive, $83.8 \pm 1.0\%$; Extended Data Fig. 4b–f), and the cells were heat-shocked¹⁴ before cryopreservation. Consistent with our previous work¹⁵, the expression of cTnT was lower in iPSC-CMs than in adult hearts (Extended Data Fig. 4g), indicative of cellular immaturity. Cells were treated with our previously reported pro-survival cocktail (PSC) before transplantation^{11,16}, and post-thawing viability as indicated by trypan blue staining was $74.5 \pm 4.1\%$. Spontaneous beating was observed *in vitro*, which was synchronous with the fluorescent transients of G-CaMP7.09 (Extended Data Fig. 3b–e, Supplementary Video 1). Consistent with previous work¹⁷, the firing rate of G-CaMP7.09 fluorescence was substantially decreased after ryanodine treatment (Extended Data Fig. 3c). Treatment with the L-type calcium-channel blocker nifedipine led to cessation of firing (Extended Data Fig. 3d), whereas the addition of caffeine—which opens sarcoplasmic reticulum ryanodine channels—promoted firing fluorescence (Extended Data Fig. 3e). Note that firing fluorescence always corresponded with contraction under all conditions (Supplementary Video 2). To exclude

¹Institute for Biomedical Sciences, Shinshu University, Matsumoto 390-8621, Japan. ²Department of Cardiovascular Medicine, Shinshu University School of Medicine, Matsumoto 390-8621, Japan. ³Department of Cardiovascular Surgery, Shinshu University School of Medicine, Matsumoto 390-8621, Japan. ⁴Department of Pediatrics, Shinshu University School of Medicine, Matsumoto 390-8621, Japan. ⁵Ina Research Inc., Ina 399-4501, Japan. ⁶Department of Molecular Life Science, Tokai University School of Medicine, Isehara 259-1193, Japan. ⁷Brain Science Institute, Saitama University, Saitama 338-8570, Japan. ⁸Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan. ⁹Chromosome Engineering Research Center, Tottori University, Yonago 683-8503, Japan. ¹⁰Institute for Integrated Cell-Material Sciences (WPI-ICeMS), Kyoto University, Kyoto 606-8501, Japan.

*These authors contributed equally to this work.

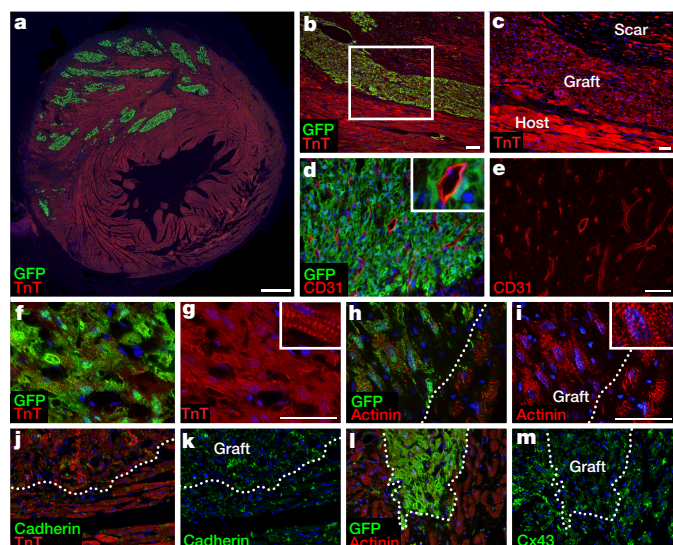


Figure 1 | Transplanted iPSC-CMs partially remuscularize infarcted cynomolgus monkey hearts. **a–m**, Fluorescence microscopic images of cynomolgus monkey hearts subjected to myocardial infarction and transplantation of iPSC-CMs. Grafts were studied on day 84 post-transplantation. **a**, A substantial number of grafted cardiomyocytes (green) survived in the anterior portion of the left ventricle. Scale bar, 1 mm. **b, c**, Grafted cells located in the border zone. Note that almost all of the GFP⁺ cells are cTnT⁺. **c**, Enlargement of the box in **b**. Scale bars, 100 μ m (**b**) and 50 μ m (**c**). **d, e**, Graft cardiomyocytes were well vascularized by the host-derived (GFP[−]/CD31⁺) endothelial cells. Representative endothelium is shown with higher magnification in the inset in **d**. Scale bar, 50 μ m. **f–i**, Graft cardiomyocytes showed sarcomeric structures identified by cTnT and α -actinin (Actinin) staining. Representative sarcomeric structures with higher magnification are shown in the insets in **g, i**. Scale bars, 50 μ m. **j–m**, Expression of cell adhesion protein pan-cadherin (Cadherin) and gap junction protein connexin 43 (Cx43) in the graft and host tissue. Scale bars, 50 μ m. **f, g, h, i, j, k**, and **l, m** denote paired images. **h–m**, Dashed lines indicate the border between graft and host tissues.

the possibility that the G-CaMP7.09 transients were generated by stretch-activated calcium intake, we cultured G-CaMP7.09-expressing iPSC-CMs on stretchable Parafilm. When the cells were treated with 40 mM 2,3-butanedione 2-monoxime (BDM), the cardiomyocytes stopped beating, whereas G-CaMP7.09 transients were sustained for a few minutes (Extended Data Fig. 3f, Supplementary Video 3). After cessation of the fluorescent transients, no G-CaMP7.09 signal was observed in response to passive stretching, but treatment with caffeine restored the G-CaMP7.09 transients (Extended Data Fig. 3g, h). These findings strongly indicate that the G-CaMP7.09 fluorescent transients in iPSC-CMs are reflective of their contraction via calcium-induced calcium release from ryanodine receptors.

We first transplanted 4×10^8 iPSC-CMs suspended in PSC into MHC-mismatched monkeys treated with methylprednisolone and tacrolimus ($n=2$) and found that grafted cardiomyocytes were thoroughly rejected as the result of severe infiltration of T lymphocytes 4 weeks after transplantation, as determined by histological analysis (Extended Data Fig. 6a, b). Subsequently, ten 4–5-year-old female cynomolgus monkeys were used as recipient animals. Five monkeys in which either of the MHC haplotypes was identical to that of the donor (HT4, Extended Data Fig. 1a) received iPSC-CMs and the other received PSC vehicle (Extended Data Fig. 5b); all animals were treated with methylprednisolone and tacrolimus. *In vitro* mixed lymphoid reactions indicated no or little immune response when HT4 homozygous cells were cocultured with HT4 heterozygous cells (Extended Data Fig. 1c). Myocardial infarction was induced by 3 h ischaemia followed by reperfusion (Supplementary Video 4). Two weeks later, either 4×10^8 iPSC-CMs suspended in PSC or vehicle alone were injected into the

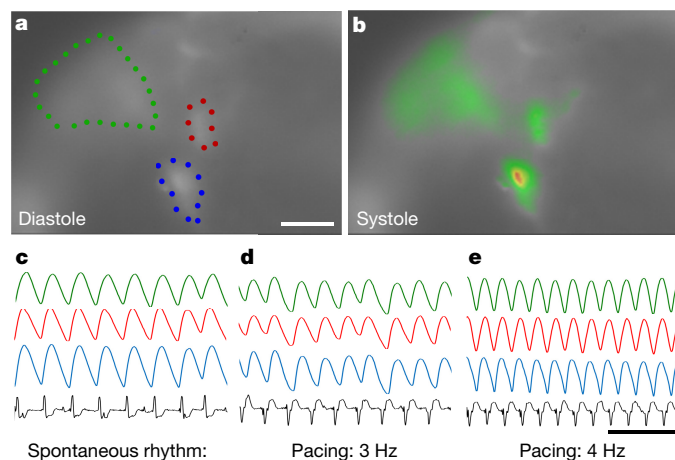


Figure 2 | iPSC-CMs electrically couple with the host heart.

a, b, Intravital fluorescence image of the Langendorff-perfused heart showing flashing (dotted lines) fluorescence of the calcium indicator, G-CaMP7.09, in the heart. Scale bar, 2 mm. **c–e**, G-CaMP7.09 fluorescent signals for green, red, and blue regions of interest indicated in **a**, as well as the ECG (black). Contractions in all three regions are synchronous with the host ECG when the heart beats spontaneously and is paced at ≤ 4 Hz (240 bpm). Scale bar, 1 s.

infarct zone and the border zone. Animals were euthanized and underwent full necropsy 12 weeks post-transplantation. None showed any macro- or microscopic tumour formation (Extended Data Fig. 7a–h). All animals showed patchy scar formation in the heart (percentage scar area/left ventricle area: $10.2 \pm 0.7\%$ (PSC vehicle) and $8.8 \pm 1.0\%$ (iPSC-CMs), $P=0.30$, Extended Data Fig. 8a). The average plasma tacrolimus trough level 12 weeks post-transplantation was 24.4 ± 3.1 ng ml^{−1} and 22.2 ± 2.2 ng ml^{−1} in PSC-vehicle and iPSC-CM recipient animals, respectively. The iPSC-CM recipients showed partial remuscularisation of the scar tissue by the grafted cardiomyocytes (percentage graft area/scar area: $16.3 \pm 5.0\%$, Fig. 1a–c and Extended Data Fig. 7i–p), which was well vascularised by the host vessels (Fig. 1d, e). More than 99% of the graft GFP⁺ cells were cTnT⁺ cardiomyocytes (Fig. 1b, c). Despite the lower overall expression of cTnT in iPSC-CMs in the graft compared to host CMs (Fig. 1c and Extended Data Fig. 4g), grafted cardiomyocytes showed clear sarcomere structure with cTnT and α -actinin (Fig. 1f–i). Grafts localized to the border zone and within the scar area (Extended Data Fig. 7i–p). Additionally, some grafts were surrounded by scar tissue and appeared to be isolated from the host myocardium (Fig. 1b, c). However when we looked at the same grafts at different levels of section, we could see that most were in direct contact with host cardiomyocytes (Extended Data Fig. 7l–n). Expression of the cell adhesion protein cadherin and the gap junction protein connexin 43 was observed in the grafts, but was relatively rare compared to that in the host (Fig. 1j–m). Immunohistochemical staining for CD45 (leukocytes), CD3 (T lymphocytes), and CD20 (B lymphocytes) on day 84 post-transplantation revealed no evidence of acute graft rejection (Extended Data Fig. 6c–i).

To confirm electrical coupling of the grafted cardiomyocytes to the host heart, all iPSC-CM-transplanted hearts were subjected to intravital G-CaMP7.09 fluorescence imaging. After deep anaesthesia, the heart was perfused with cold cardioplegia solution, excised, and transported to a Langendorff setup. Next, it was reperfused with oxygenated Tyrode's solution and resumed spontaneous beating, and G-CaMP signalling and electrocardiograms (ECGs) were recorded. We observed multiple flashing fluorescent signals that were synchronous with each other and with the host ECG in all five animals (Fig. 2a–e, Supplementary Video 5, and Extended Data Fig. 8a). Grafted cardiomyocytes showed 1:1 coupling with host cardiomyocytes (Fig. 2c–e), but graft activation was delayed when the hearts were paced at 4 Hz

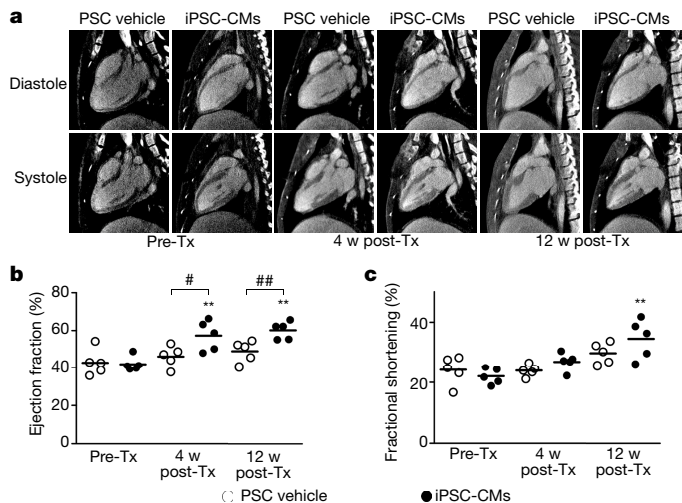


Figure 3 | Transplantation of iPSC-CMs improves cardiac contractile function. Cardiac contractile function was evaluated before transplantation (Pre-Tx) and at 4 weeks post-transplantation (4 w post-Tx) and 12 weeks post-transplantation (12 w post-Tx) by μ CT and echocardiography. **a**, Representative longitudinal axis diastolic and systolic μ CT images. **b**, Ejection fraction as assessed by μ CT. **c**, Fractional shortening as assessed by echocardiography. $n = 5$ in each group. $^*P < 0.05$, $^{**}P < 0.01$ between vehicle and iPSC-CM; $^{***}P < 0.01$ versus Pre-Tx.

(Extended Data Fig. 9a, b). This slower propagation probably reflects the limited formation of gap junctions in the graft tissue (Fig. 1m). To evaluate cardiac contraction, we used a novel micro-computed tomography (μ CT) system (Fig. 3a), as well as echocardiography. Contractile function analysis by μ CT in intact cynomolgus monkeys ($n = 5$) revealed consistent ejection fractions ($64.6 \pm 1.5\%$). Echocardiography and μ CT were performed on days -2 , 28 and 84 relative to transplantation (Fig. 3a–c and Extended Data Fig. 10a–d). Echocardiography-based analysis of fractional shortening revealed that shortening in iPSC-CM recipients tended to be higher, albeit not significantly, than in vehicle-treated recipients on days 28 and 84. Fractional shortening in transplanted animals improved significantly only on day 84 compared to day -2 (Fig. 3c). μ CT analysis seemed to be superior to echocardiography for evaluating apical cardiac contraction (Supplementary Video 6) because contractile function of the apex was generally difficult to evaluate by echocardiography after sternotomy, owing to post-operative adhesion. Furthermore, the ejection fraction was significantly higher

in the iPSC-CM recipients than in vehicle-treated animals on days 28 and 84 and compared to pre-transplantation readings as analysed by μ CT (Fig. 3b). There was also a reasonable negative correlation between scar size and ejection fraction ($r = -0.67$, Extended Data Fig. 8b) and a positive correlation between graft size and ejection fraction ($r = 0.91$, Extended Data Fig. 8c). In vehicle-treated animals, the plasma B-type natriuretic peptide (BNP) level was highest on day 0 and gradually declined throughout the experimental period (the BNP levels on days 56 and 84 were significantly lower than on day 0); although the levels in iPSC-CM recipients tended to increase on day 28 when compared to day 0, they decreased on days 56 and 84. However, BNP levels did not differ significantly between the two groups at any time point (Extended Data Fig. 10e).

To assess the electrophysiological consequences of iPSC-CM grafting, the animals were subjected to Holter ECG monitoring on days -2 , 7, 14, 28, 42, 56, 70 and 80 relative to the day of transplantation. No ventricular tachycardia was observed before transplantation in either iPSC-CM or vehicle recipients. Episodes of sustained ventricular tachycardia (Fig. 4a) were observed only after transplantation. The duration of the sustained ventricular tachycardia peaked at day 14 and shortened considerably thereafter in 4 out of 5 animals (Fig. 4b and Extended Data Fig. 8d). Similarly, the incidence of any (sustained or non-sustained) ventricular tachycardia (Fig. 4c and Extended Data Fig. 9c–f) and sustained ventricular tachycardia (Fig. 4d) peaked on day 14 and declined gradually throughout the rest of the study period. Notably, while all of the recipients of iPSC-CMs showed sustained ventricular tachycardia on day 14, none was apparent on days 56 and 84 (Fig. 4d). The fastest observed ventricular tachycardia exceeded 240 bpm (beats per minute) (Extended Data Fig. 8d), but none of the animals showed any abnormal behaviour, such as syncope, throughout the study period.

This study demonstrated that allogeneic transplantation of MHC-matched iPSC-CMs can provide long-term graft survival in the infarcted hearts of non-human primates. Both MHC and minor antigens^{18–20} have been shown to play important roles in the immune response following allogeneic transplantation. In fact, one research group recently observed graft rejection following MHC-matched iPSC-CM transplantation into the subcutaneous tissue of allogeneic cynomolgus monkeys²¹. Given that grafted cardiomyocytes survived for 12 weeks without immune rejection in all five iPSC-CM recipients in this study, it is reasonable to assume that a combination of methylprednisolone and tacrolimus is sufficient to prevent immune rejection of transplanted allogeneic cardiomyocytes. Nevertheless, further studies are required to establish the minimum amount of immunosuppression required to control immune rejection following cell transplantation.

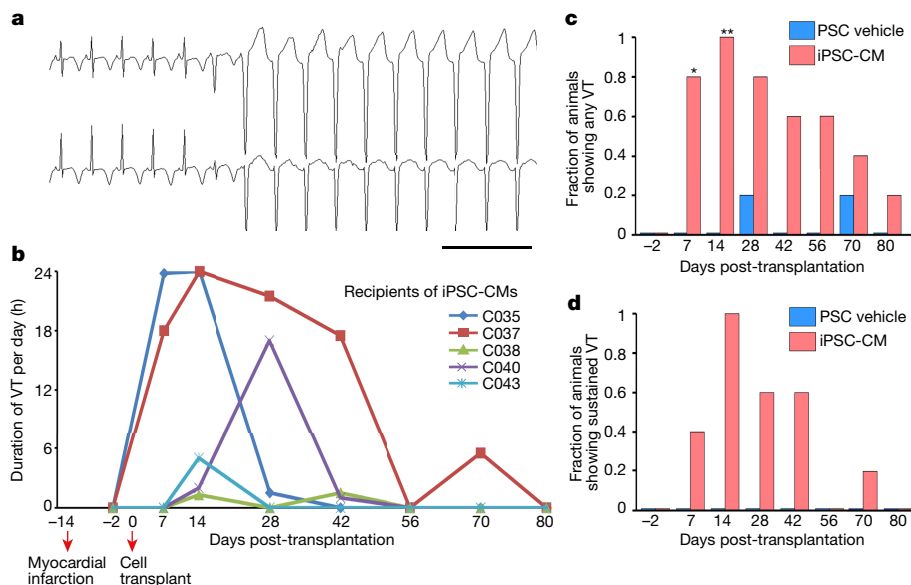


Figure 4 | Electrical consequences of transplantation of iPSC-CMs.

a, Representative traces of ventricular tachycardia (VT) in a recipient of iPSC-CMs. Scale bar, 1 s. **b**, Duration of sustained ventricular tachycardia in iPSC-CM recipients. Note that none of the recipients of PSC vehicle showed sustained ventricular tachycardia throughout the study period. **c**, Fraction of animals showing any (sustained or non-sustained) ventricular tachycardia. **d**, Fraction of animals showing sustained ventricular tachycardia. $n = 5$ in each group. $^*P < 0.05$; $^{**}P < 0.01$ versus PSC vehicle.

We also demonstrated that iPSC-CMs integrated and improved cardiac contractile function in a primate infarct model. Demonstration of electrically coupled grafts strongly supports the notion that grafted cardiomyocytes improve cardiac contractile function by creating new force-generating units, although other mechanisms—such as paracrine effects—cannot be excluded.

Although our previous study showed that transplantation of human cardiomyocytes suppressed ventricular arrhythmias in injured guinea-pig hearts⁹, Chong *et al.* observed post-transplant arrhythmias in the non-human primate heart after human ES-cell-derived cardiomyocyte transplantation²². The most likely reason for these inconsistencies is that the non-human primate heart is more similar to the human heart than the hearts of rodents with respect to size and beating rate. It is noteworthy that allogeneic transplantation in large animal models, as is the case in clinical studies, is thought to be the most sensitive model to detect post-transplant arrhythmias. In fact, although idioventricular rhythms represented the majority of sustained ventricular arrhythmias observed in the xenogeneic transplantation study²², our study demonstrated that allogeneic transplantation of iPSC-CMs significantly increased the incidence of ventricular tachycardia. The incidence of sustained ventricular tachycardia did not seem to correlate with graft size, coupled graft number, scar size or cardiac contractile function in our study (Extended Data Fig. 8a, d). The reduced incidence of ventricular tachycardia over time may reflect the reduced portion of grafted cardiomyocytes; however, pluripotent stem cell-derived cardiomyocytes have been shown to proliferate *in vivo* and the graft area grew until 4 weeks post-transplantation¹⁴. A more likely possibility is that the decreased incidence of ventricular tachycardia resulted from maturation of graft CMs *in vivo*²². Notably, all iPSC-CM recipients survived for 12 weeks until the end of the study without any abnormal behaviour, and post-transplant arrhythmias seemed to be transient, peaking on day 14 post-transplantation and gradually decreasing thereafter. These results suggest that iPSC-CM transplant-induced ventricular tachycardia was non-lethal and transient.

There are a few limitations in the study design. First, we tested only one iPS cell line, so additional studies using multiple cell lines will be required. Second, our animal model is not always clinically relevant in view of its relatively small infarct size in young adolescent monkeys. Finally, the 12-week observation period after cell transplantation does not allow a definitive conclusion regarding graft survival without chronic rejection and a longer follow-up study will be required to investigate chronic rejection and the risks associated with immunosuppressant use further.

In conclusion, allogeneic transplantation of iPSC-CMs led to integrated graft survival and improved cardiac contractility for at least 12 weeks in a non-human primate myocardial infarction model. Transient, non-lethal ventricular tachycardia was significantly increased by the iPSC-CMs, generating the need for more effort to control arrhythmias before clinical application.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 October 2015; accepted 31 August 2016.

Published online 10 October 2016.

- Deleidi, M., Hargus, G., Hallett, P., Osborn, T. & Isacson, O. Development of histocompatible primate-induced pluripotent stem cells for neural transplantation. *Stem Cells* **29**, 1052–1063 (2011).
- Laflamme, M. A. & Murry, C. E. Heart regeneration. *Nature* **473**, 326–335 (2011).
- Lalit, P. A., Hei, D. J., Raval, A. N. & Kamp, T. J. Induced pluripotent stem cells for post-myocardial infarction repair: remarkable opportunities and challenges. *Circ. Res.* **114**, 1328–1345 (2014).
- Shiba, Y., Hauch, K. D. & Laflamme, M. A. Cardiac applications for human pluripotent stem cells. *Curr. Pharm. Des.* **15**, 2791–2806 (2009).
- Bach, F. H., Bach, M. L. & Sondel, P. M. Differential function of major histocompatibility complex antigens in T-lymphocyte activation. *Nature* **259**, 273–281 (1976).

- Petersdorf, E. W. The major histocompatibility complex: a model for understanding graft-versus-host disease. *Blood* **122**, 1863–1872 (2013).
- Shiina, T. *et al.* Discovery of novel MHC-class I alleles and haplotypes in Filipino cynomolgus macaques (*Macaca fascicularis*) by pyrosequencing and Sanger sequencing: Mafa-class I polymorphism. *Immunogenetics* **67**, 563–578 (2015).
- Blancher, A. *et al.* Study of MHC class II region polymorphism in the Filipino cynomolgus macaque population. *Immunogenetics* **66**, 219–230 (2014).
- Shiba, Y. *et al.* Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts. *Nature* **489**, 322–325 (2012).
- Shiba, Y. *et al.* Electrical integration of human embryonic stem cell-derived cardiomyocytes in a guinea pig chronic infarct model. *J. Cardiovasc. Pharmacol. Ther.* **19**, 368–381 (2014).
- Laflamme, M. A. *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat. Biotechnol.* **25**, 1015–1024 (2007).
- Zhang, J. *et al.* Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method. *Circ. Res.* **111**, 1125–1136 (2012).
- Tohyama, S. *et al.* Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell* **12**, 127–137 (2013).
- Laflamme, M. A. *et al.* Formation of human myocardium in the rat heart from human embryonic stem cells. *Am. J. Pathol.* **167**, 663–671 (2005).
- Minami, I. *et al.* A small molecule that promotes cardiac differentiation of human pluripotent stem cells under defined, cytokine- and xeno-free conditions. *Cell Reports* **2**, 1448–1460 (2012).
- Gautam, M. *et al.* Transplantation of adipose tissue-derived stem cells improves cardiac contractile function and electrical stability in a rat myocardial infarction model. *J. Mol. Cell. Cardiol.* **81**, 139–149 (2015).
- Li, J., Qu, J. & Nathan, R. D. Ionic basis of ryanodine's negative chronotropic effect on pacemaker cells isolated from the sinoatrial node. *Am. J. Physiol.* **273**, H2481–H2489 (1997).
- Derks, R. A., Jankowska-Gan, E., Xu, Q. & Burlingham, W. J. Dendritic cell type determines the mechanism of bystander suppression by adoptive T regulatory cells specific for the minor antigen HA-1. *J. Immunol.* **179**, 3443–3451 (2007).
- Kwon, J. *et al.* Impact of leukocyte function-associated antigen-1 blockade on endogenous allospecific T cells to multiple minor histocompatibility antigen mismatched cardiac allograft. *Transplantation* **99**, 2485–2493 (2015).
- Vokaer, B. *et al.* Critical role of regulatory T cells in Th17-mediated minor antigen-disparate rejection. *J. Immunol.* **185**, 3417–3425 (2010).
- Kawamura, T. *et al.* Cardiomyocytes derived from MHC-homozygous induced pluripotent stem cells exhibit reduced allogeneic immunogenicity in MHC-matched non-human primates. *Stem Cell Reports* **6**, 312–320 (2016).
- Chong, J. J. *et al.* Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature* **510**, 273–277 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Ichihara, N. Ishimine, Y. Karatsu, the Rigaku Corporation, Brainvision Inc. and the Keyence Corporation for assistance with the experiments and Astellas Pharma Inc. for the gift of tacrolimus. We also appreciate the scientific advice of M. A. Laflamme, J. Chong and N. Saito. This work was supported by research grants (to Y.S.) from the Japan Society for the Promotion of Science KAKENHI (grant no. 26293182), Japan Agency for Medical Research and Development, Takeda Science Foundation, Astellas Foundation for Research on Metabolic Disorders, Mochida Memorial Foundation for Medical and Pharmaceutical Research, and Japan Heart Association. The experiments to generate the G-CaMP7.09 plasmid were supported by grants from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to J.N. (grant no. 26115504) and M.O. (grant no. 25116504) and a grant from the Regional Innovation Cluster Program (City Area Type, Central Saitama Area) to J.N.

Author Contributions Y.S. designed the study. Y.S., T.G., T.Se., Y.W., H.I., Y.T., K.S. and D.I. performed all animal procedures. T.O., N.S. and Y.S. performed histological analysis. K.S. and D.I. analysed Holter ECGs. N.U., Y.K., and M.O.s. performed karyotype analysis of iPSCs. T.Sh. analysed RNA sequences of cynomolgus MHC. K.O. and U.I. analysed all other data and provided administrative assistance. M.Oh. and J.N. generated the G-CaMP7.09 plasmid. *In vitro* fluorescent imaging studies were performed by I.M. The manuscript was written by Y.S., T.Sh., M.Oh., I.M. and N.U.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.S. (yshiba@shinshu-u.ac.jp).

Reviewer Information Nature thanks T. Braun, K. Fukuda, T. Kamp and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Screening for cynomolgus monkeys with MHC homologous and heterozygous haplotypes. Total RNA was isolated from peripheral white blood cells using TRIzol reagent (Thermo Fisher Scientific) and treated with DNase I (Thermo Fisher Scientific). cDNA was synthesized using ReverTra Ace (Toyobo) with oligo d(T) primer. A set of previously reported cynomolgus macaque MHC (*Mafa*) class I-specific primers was used for RT-PCR amplification⁷, and *Mafa-DRB*, *Mafa-DQA1*, *Mafa-DQB1*, *Mafa-DPA1*, and *Mafa-DPB1* were amplified with the following specific primers: *Mafa-DRB* (DRB_PHI_F1, 5'-GCTCCCTGGAGGCTCCTG-3'; DRB_PHI_R1-1, 5'-ACCAGGAGGGTGTGGTGC-3'; DRB_PHI_R1-2, 5'-ACCAGCAGGGTGTGGTGC-3'; DRB_PHI_R1-3, 5'-ACCAGGAGGTTGTGGTGC-3' and DRB_PHI_R1-4, 5'-ACCAGGAGGCTGTGGTGC-3'), *Mafa-DQA1* (DQA_PHI_F1, 5'-ATCCTAAACAAAGCTCTG-3' and DQA_PHI_R2, 5'-TGTGATGTTTACCACAGG-3'), *Mafa-DQB1* (DQB_PHI_F1, 5'-CTGTGACCTTGATGCTGG-3' and DQB_PHI_R1, 5'-AGACCAGCAGGTTGTGGT-3'), *Mafa-DPA1* (DPA_PHI_F1_1, 5'-ATGTTCCAGACCAGAGCT-3'; 5'-ATGTTCCAGACCAGAGCT-3' and DPA_PHI_R1, 5'-TTGTCAATGTGGCAGATG-3') and *Mafa-DPB1* (DPB_PHI_F2, 5'-GCCACTCCAGAGAATAC-3' and DPB_PHI_R2, 5'-GAGCAGGTTGTGGTGTGCTG-3').

In addition, we designed MHC-specific fusion primers containing the Roche 454 titanium adaptor (A in forward and B in reverse primer) and a 10-bp multiple identifier (MID). In brief, each 20- μ l RT-PCR mixture contained 10 ng cDNA, 0.4 U high-fidelity KOD FX polymerase (Toyobo), 2 \times PCR buffer, each dNTP at 2 mM, and each primer at 0.5 μ M. The thermal cycling program was as follows: 25 cycles at 98 °C for 10 s, 58 °C for 30 s, and 68 °C for 30 s. The PCR products were purified with the QIAquick PCR purification kit (Qiagen) and quantified by the picogreen assay (Invitrogen) in a fluoroskan ascent microplate fluorometer (Thermo Fisher Scientific). The PCR products were mixed at equimolar concentrations and then diluted according to the manufacturer's recommendations (Roche). Emulsion PCR, breaking and bead enrichment and deposition into a picotiterplate were performed according to the manufacturer's protocol (Roche). Image processing, signal correction and base calling were performed using the GS run processor version 3.0 (Roche) with full processing for shotgun or paired-end filter analysis. Quality-filtered sequence reads that passed the assembler software (single sff file) were binned into separate sequence sff files on the basis of the MID labels, using the sff file software (Roche). These files were further trimmed to remove poor-quality sequence at the end of the reads with quality values <20. The *Mafa* class I and class II alleles were assigned by matching the sequence reads with all known *Mafa* class I allele sequences in the IMGT/MHC-NHP database²³ with GS reference mapper version 3.0, using the following parameter settings: 99% and 100% matching (for class I and class II alleles, respectively), minimum overlap length of 200 and alignment identity score of 10. We selected HT4 homozygous and heterozygous animals that had the following *Mafa* class I and class II alleles: *Mafa-A1*089:03*, *Mafa-A2*05:50*, *Mafa-A3*13:03:01*, *Mafa-B*046:01:02*, *Mafa-B*050:08*, *Mafa-B*057:04*, *Mafa-B*060:02*, *Mafa-B*072:01*, *Mafa-B*104:03*, *Mafa-B*114:02*, *Mafa-B*144:03N*, *Mafa-I*01:12:01*, *Mafa-DRB1*03:21*, *Mafa-DRB1*10:07*, *Mafa-DQA1*01:07:01*, *Mafa-DQB1*06:08*, *Mafa-DPA1*02:05*, and *Mafa-DPB1*15:04* (Extended Data Fig. 1a). Since only 0.84% of animals showed the HT4 haplotype in either MHC, we designated HT4 heterozygous monkeys as recipients of iPSC-CMs and non-HT4 monkeys as recipients of PSC vehicle (no randomization between groups). All experiments and analyses were performed under blinded conditions. No statistical methods were used to predetermine sample size.

Generation of a G-CaMP7.09-reporter cynomolgus iPSC line. Skin fibroblasts were isolated from a male MHC-homozygous cynomolgus monkey. The fibroblasts were transfected with a combination of plasmid vectors encoding *OCT4*, *SOX2*, *KLF4*, and *L-MYC* as described previously²⁴. The cynomolgus iPSCs were maintained on SNL feeder cells (Cell Biolabs) treated with mitomycin C (Sigma-Aldrich) in essential 8 medium (Thermo Fisher Scientific). We developed a novel fluorescent calcium indicator G-CaMP7.09. Briefly, a cDNA encoding G-CaMP7.09 was constructed by replacing the 1.13-kb *SacI*/*Clal* fragment of G-CaMP7 cDNA with the corresponding 1.13-kb fragment of G-CaMP5.09 cDNA²⁵. G-CaMP7.09 differs from G-CaMP7 by an N205S mutation in the circularly permuted, enhanced GFP (EGFP) domain and an L36M mutation in the calmodulin (CaM) domain (Extended Data Fig. 3a). The G-CaMP7.09 cDNA was subcloned into a pEGFP-N1 vector (Clontech) with a CMV promoter as described previously²⁵, for expression in cynomolgus iPSCs.

The G-CaMP7.09 plasmid was electroporated into cynomolgus iPSCs cultured in essential 8 medium supplemented with 10 mM Y-27632 (Thermo Fisher Scientific). The electroporation conditions were 1400 V pulse voltage, 10 ms pulse width and 2 pulses. Transfected cells were selected with 100 μ g ml⁻¹ G418 for 7 days. Successful transfection was confirmed by identification of green fluorescence by flow cytometry.

Karyotype analysis of cynomolgus iPSCs. The iPSCs were treated with 0.025 μ g ml⁻¹ colcemid for 4 h. The cells were collected by trypsinisation, incubated in 0.075 M KCl for 15 min, and fixed with methanol and acetic acid (3:1). Then, the chromosomes were spread on slides. The chromosome spreads were stained with quinacrine mustard and Hoechst33258 to enumerate chromosomes, following a standard protocol. Images were captured using an Axio ImagerZ2 fluorescence microscope (Carl Zeiss GmbH).

RT-PCR analysis of cynomolgus iPSCs and iPSC-CMs. Total RNA was isolated from cynomolgus iPSCs, iPSC-CMs or adult heart, using an RNeasy mini kit (Qiagen). cDNA was synthesized from 1 μ g of total RNA with superscript III (Invitrogen) according to the manufacturer's instructions. The cDNA was PCR-amplified using the following primers:

OCT4, 5'-CAGATCAGCCACATTGCCAG-3' and 5'-CAAAAGCCCC TGGCACAACCTCT-3'; *NANOG*, 5'-CCTATGCCTGTGATTTGTGGG-3' and 5'-AGGTTGTTTGCCTTTGGGAC-3'; *SOX2*, 5'-GGTTACCTCTTCC TCCACTCC-3' and 5'-CCTCCCATTTCCCTCGTTT-3'; *TNNT2*, 5'-AAGG AAGCTGAAGATGGCCC-3' and 5'-GGGCTGCTTCTGGATGTAA-3';

GAPDH, 5'-AATCCCATCACCATTCCAGGAG-3' and 5'-CACCTGTG TGTGTAGCCAAATTC-3'.

The thermal cycling conditions were as follows: denaturation at 94 °C for 30 s, 30 cycles of 10 s at 98 °C; 30 s at 55 °C for *GAPDH*, 60 °C for *OCT4*, *NANOG*, and *SOX2*; and 30 s at 68 °C; final extension at 72 °C for 1 min.

Immunohistochemical analysis of cynomolgus iPSCs for pluripotency markers. Cells were fixed with 2% paraformaldehyde for 10 min and stained with antibodies against OCT4 (clone: c-10), NANOG (rabbit polyclonal) and SSEA4 (clone: MC-813-70), followed by goat anti-mouse-594 or goat anti-rabbit-488 (Thermo Fisher Scientific).

Teratoma formation assay. Undifferentiated iPSCs (10⁷) in PBS were injected into the adductor longus muscle of male Fox Chase SCID mice (Charles River). When subcutaneous tumours were apparent at the site of transplantation (typically 6 weeks post-transplantation), the mice were euthanized and the tumours were excised and fixed with 4% paraformaldehyde.

Mixed lymphoid reaction. *In vitro* mixed lymphoid reaction was performed as described previously²⁶ with modifications. Prior to enrolment, a 7-ml blood sample was collected from each animal via venous puncture of the femoral vein. Peripheral blood mononuclear cells (PBMCs) were isolated using a vacutainer cell preparation tube (BD Biosciences) according to manufacturer's instructions. Recipient animal-derived PBMCs (10⁶) were co-cultured for 5 days with the same number of donor-derived PBMCs pre-treated with 25 μ g ml⁻¹ mitomycin C. The cellular proliferation was monitored using a 5-bromo-2'-deoxy-uridine (BrdU) labelling and detection kit (Roche) according to the manufacturer's instructions. The control sample consisted of recipient animal-derived PBMCs without donor-derived cells. BrdU incorporation was expressed relative to the control.

iPSC-CM preparation. Undifferentiated cynomolgus iPSCs were differentiated into iPSC-CMs using the matrix sandwich method¹². Briefly, undifferentiated iPSCs were plated on a Matrigel-coated culture dish (Corning) and cultured in essential 8 medium for a few days. When the cells reached 80–90% confluency, Matrigel was added to the medium, and the cells were treated with activin A (R&D) and subsequently, bone morphogenic protein 4 (BMP4; R&D). On day 14 after differentiation, the cells were exposed to glucose-free medium for 3 days to enrich cardiomyocytes¹³. The cells were heat-shocked and cryopreserved. Cardiac purity was determined by immunostaining of cTnT (clone CT3), followed by anti-mouse IgG1 conjugated with phycoerythrin, using a FACSCanto II (BD Biosciences). G-CaMP7.09 fluorescence was measured using a FACSCanto II. Before transplantation, 4 \times 10⁸ cells were thawed and processed using a previously reported pro-survival cocktail¹¹. Cell viability was determined by a trypan blue assay.

Animal surgeries. Based on the national regulations and guidelines, all experimental procedures were reviewed by the Committee for Animal Experiments and finally approved by the president of Shinshu University and Ina Research.

For major surgeries, the animals were anaesthetized by intra-muscular injection of ketamine and xylazine, intubated with a tracheal tube (4-mm diameter), and ventilated with 1.5% isoflurane. Buprenorphine was routinely administered subcutaneously to provide post-operative pain relief. Blood pressure, ECG and oxygen saturation were monitored during surgery. Phenylephrine was administered intravenously to maintain appropriate blood pressure. After median sternotomy, a 4-0 silk suture was passed through the myocardium at the mid-left anterior descending (LAD) coronary artery and threaded through a polyethylene tube (SP110, Natsume). A silicon tube was placed on top of the polyethylene tube and was tied off with a suture (Supplementary Video 4). Before induction of myocardial infarction, 1 mg kg⁻¹ lidocaine and 1000 U heparin were administered intravenously. The same dose of heparin was subsequently administered every h until reperfusion. After 3 h of occlusion of the mid LAD, the heart was reperfused by removing the tubing. Immune suppression was achieved by daily

intra-muscular injection of methylprednisolone and tacrolimus (Astellas Pharma Inc.). Methylprednisolone was administered at $10 \text{ mg kg}^{-1} \text{ day}^{-1}$ from the day before transplantation for 3 days and at $1 \text{ mg kg}^{-1} \text{ day}^{-1}$ thereafter. Tacrolimus was administered at $0.1 \text{ mg kg}^{-1} \text{ day}^{-1}$, 2 days before transplantation.

On day 14 after myocardial infarction, the animals underwent a second sternotomy and the heart was exposed. Either iPSC-CMs (4×10^8) suspended in PSC or PSC vehicle were delivered intra-myocardially into the infarct and border zones via 10 injections of $100 \mu\text{l}$ each using a 29-gauge injection needle.

Echocardiography. Echocardiography was performed on days -2, 28 and 84 relative to transplantation. After intra-muscular injection of the ketamine and xylazine anaesthetic mixture, the left-ventricular end-diastolic dimension (LVEDD), left-ventricular end-systolic dimension (LVESD) and heart rate were measured by transthoracic echocardiography (GE Vivid7) with a 10-MHz paediatric transducer. Fractional shortening (FS) was calculated using the following equation: $\text{FS} = 100 \times ((\text{LVEDD} - \text{LVESD}) / \text{LVEDD})$. All measurements were taken over three consecutive cardiac cycles and averaged. An operator who was blinded to the study groups performed all measurements.

MicroCT. MicroCT was performed on the same days as echocardiography. Anaesthetized animals were intubated and mechanically ventilated with 1.5% isoflurane. Radiopaque agent (Iopamiron Inj., Bayer) was infused at 8 ml min^{-1} . The hearts were imaged in an R_mCT AX (Rigaku, Japan), using the following settings: 80 kV; 500 μA ; field of view, 100 mm. Motion cycles of cardiac contraction and ventilation were automatically synchronised by the μCT system. Left ventricular end-diastolic volume (LVEDV) and end-systolic volume (LVESV) were measured using Ziostation2 software (Amin). Left ventricular ejection fraction (LVEF) was calculated using this equation: $\text{LVEF} (\%) = 100 \times ((\text{LVEDV} - \text{LVESV}) / \text{LVEDV})$. An operator who was blinded to the study groups performed all measurements.

Holter ECG. The Holter ECG recordings were performed on days -2, 7, 14, 28, 42, 56, 70 and 80 relative to transplantation. The area intended for electrode placement was prepared by shaving. The electrodes were placed in a 2-lead precordial system and connected to a Holter recorder. A jacket was placed on the animal to protect the ECG system and a 24-h ECG was recorded. Ventricular tachycardia was defined as four or more consecutive premature ventricular complexes with a ventricular rate faster than 180 bpm. Sustained ventricular tachycardia was defined as ventricular tachycardia sustained longer than 30 s. An operator who was blinded to the study groups performed all analyses.

Blood test. Peripheral blood was collected on days 0, 28, 56 and 84 relative to transplantation. Plasma was isolated to measure BNP levels, using an automatic immunoassay kit (TOSOH). Whole blood was used to measure trough levels of tacrolimus by electrochemiluminescence immunoassay (SRL).

Fluorescence imaging of G-CaMP7.09-expressing iPSC cell-derived cardiomyocytes. For *in vitro* fluorescent imaging, G-CaMP7.09-expressing cardiomyocytes were cultured on Matrigel-coated culture dishes (5×10^5 cells per cm^2) or stretchable parafilm (Bemis) for 7 days in 10% FBS (GIBCO)-containing IMDM (Sigma) with 1% MEM nonessential amino acid solution (Sigma) and 2 mM L-glutamine (Sigma). A fluorescence microscope (Olympus) was used to measure the fluorescence intensity of G-CaMP7.09. Ryanodine (50 μM), caffeine (5 mM, Wako), nifedipine (50 μM , Sigma) and BDM (20 mM, Wako) were added directly into the culture medium. Intravital imaging of the hearts grafted with G-CaMP7.

09-expressing iPSC-CMs was performed on day 84 post-transplantation. After deep anaesthesia, iPSC-CM-transplanted hearts were injected with cold cardioplegia solution through the aorta (Miotect; Mochida Pharmaceutical co.) and isolated. The hearts were transferred from the animal facility to the laboratory equipped with Langendorff setup. The hearts were reperfused at 37°C with Tyrode's solution (containing $140 \text{ mmol l}^{-1} \text{ NaCl}$, $1.8 \text{ mmol l}^{-1} \text{ CaCl}_2$, $5.4 \text{ mmol l}^{-1} \text{ KCl}$, $1 \text{ mmol l}^{-1} \text{ MgCl}_2$, 11 mmol l^{-1} glucose, 5 mmol l^{-1} HEPES; bubbled with oxygen; pH 7.4). ECG (3 leads) and perfusion pressure was monitored continuously (PowerLab; ADInstruments). Epicardial G-CaMP7.09 signalling was optically recorded using a CCD camera (MiCAM02, Brainvision) through a band-pass filter (500–550 nm) when the heart was beating spontaneously or was paced at 3–4 Hz. To minimise motion artefacts, Tyrode's solution was supplemented with 15 mM BDM.

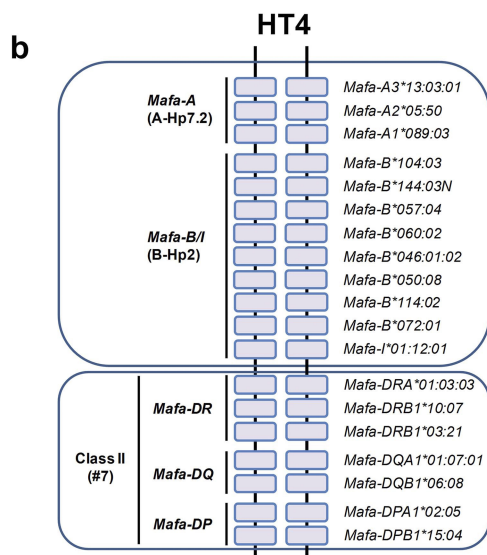
Histological analysis. On day 84 post-transplantation, the hearts were sectioned at 5-mm thickness and fixed with 4% paraformaldehyde. All sections were routinely stained with haematoxylin and eosin (HE) and picrosirius red to determine the scar region. Scar area was calculated by subtracting graft area from all fibrous areas (shown in red by picrosirius red staining) if the grafts were located in the scar (Extended Data Fig. 7i–k). The sections were immunohistologically analysed using primary antibodies against GFP (Novus, rabbit polyclonal), cTnT (clone: CT3), connexin 43 (Cx43, Abcam, rabbit polyclonal), CD45 (clone: 2B11&PD7/26/16), CD3 (clone: CD3-12), CD20 (clone: L26), CD31 (clone: JC/70A), pan cadherin (clone: CH-19), and sarcomeric α -actinin (clone: EA-53) followed by species-specific fluorescent (Molecular Probes) or biotin-conjugated (Vector Laboratories) secondary antibodies. For chromogenic detection, we used an HRP-conjugated streptavidin ABC kit (Vector Laboratories), followed by a DAB substrate kit (Vector Laboratories). Stained sections were imaged using a NanoZoomer 2.0-RS (Hamamatsu) or a Pulse-SIM BZ-X700 microscope (Keyence).

Statistical analysis. Ultrasound cardiography (UCG), μCT and BNP outcomes were analysed by an analysis of variance (ANOVA), followed by post-hoc comparisons between time points by Tukey's multiple comparison test, and unpaired *t*-test analysis to compare groups at each time point. For comparisons of the fraction of animals showing ventricular tachycardia, we used a two-sided Fisher's exact test. The percentage of cTnT and BrdU incorporation was analysed by ANOVA followed by post-hoc Tukey's multiple comparison tests. Correlations between ejection fraction and scar size or graft size were demonstrated by Pearson analysis. All statistical analyses were performed using GraphPad Prism, with the threshold for significance set at $P < 0.05$.

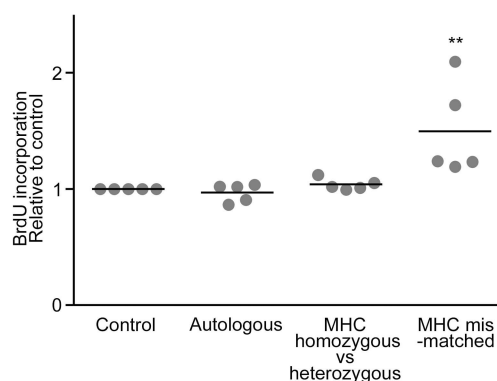
23. Robinson, J., Halliwell, J. A., McWilliam, H., Lopez, R. & Marsh, S. G. IPD--the immuno polymorphism database. *Nucleic Acids Res.* **41**, D1234–D1240 (2013).
24. Okita, K. *et al.* An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells* **31**, 458–466 (2013).
25. Ohkura, M. *et al.* Genetically encoded green fluorescent Ca^{2+} indicators with improved detectability for neuronal Ca^{2+} signals. *PLoS One* **7**, e51286 (2012).
26. Bigaud, M., Maurer, C., Vedrine, C., Puissant, B. & Blancher, A. A simple method to optimize peripheral blood mononuclear cell preparation from cynomolgus monkeys and improve mixed lymphocyte reactions. *J. Pharmacol. Toxicol. Methods* **50**, 153–159 (2004).

a

		DrpZ5-32B-C (IPS cell donor)		C035		C037		C038		C040		C043	
	Country	Philippines		Philippines		Philippines		Philippines		Philippines		Philippines	
	Sex	Male		Female		Female		Female		Female		Female	
Mafa Class I (MHC Class I)	Mafa-A	A1*089:03	A1*089:03	A1*089:03	A1*052:02	A1*089:03	A1*093:01	A1*089:03	A1*052:02	A1*089:03	A1*093:01	A1*089:03	A1*052:02
		A2*05:50	A2*05:50	A2*05:50	A4*01:04	A2*05:50	A1*074:02	A2*05:50	A4*01:04	A2*05:50	A1*074:02	A2*05:50	A4*01:04
		A3*13:03:01	A3*13:03:01	A3*13:03:01		A3*13:03:01	A8*01:01	A3*13:03:01		A3*13:03:01		A3*13:03:01	
	Mafa-B	B*104:03	B*104:03	B*104:03	B*095:01	B*104:03	B*095:01	B*104:03	B*095:01	B*104:03	B*041:01	B*104:03	B*095:01
		B*144:03N	B*144:03N	B*144:03N	B*033:02	B*144:03N	B*033:02	B*144:03N	B*033:02	B*144:03N	B*101:01:02	B*144:03N	B*033:02
		B*057:04	B*057:04	B*057:04	B*098:10	B*057:04	B*098:10	B*057:04	B*098:10	B*057:04	B*098:08	B*057:04	
		B*060:02	B*060:02	B*060:02		B*060:02		B*060:02		B*060:02		B*060:02	
		B*046:01:02	B*046:01:02	B*046:01:02		B*046:01:02		B*046:01:02		B*046:01:02		B*046:01:02	
		B*050:08	B*050:08	B*050:08		B*050:08		B*050:08		B*050:08		B*050:08	
		B*114:02	B*114:02	B*114:02		B*114:02		B*114:02		B*114:02		B*114:02	
		B*072:01	B*072:01	B*072:01		B*072:01		B*072:01		B*072:01		B*072:01	
	Mafa-I	I*01:12:01	I*01:12:01	I*01:12:01	I*01:11	I*01:12:01	I*01:11	I*01:12:01	I*01:11	I*01:12:01	I*01:11	I*01:12:01	I*01:11
Mafa Class II (MHC Class II)	Mafa-DRB	DRB1*03:21	DRB1*03:21	DRB1*03:21	DRB1*03:21	DRB1*03:21	DRB1*03:21	DRB1*03:21	DRB*W1:08	DRB1*03:21	DRB*W1:08	DRB1*03:21	DRB1*03:21
		DRB1*10:07	DRB1*10:07	DRB1*10:07	DRB1*10:07	DRB1*10:07	DRB1*10:07	DRB1*10:07	DRB*W36:01	DRB1*10:07	DRB*W36:01	DRB1*10:07	DRB1*10:07
	Mafa-DQA1	DQA1*01:07:01	DQA1*01:07:01	DQA1*01:07:01	DQA1*01:07:01	DQA1*01:07:01	DQA1*01:07:01	DQA1*01:07:01	DRB*W3:01	DQA1*01:07:01	DRB*W3:01	DQA1*01:07:01	DQA1*01:07:01
		DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03	DQA1*26:03
	Mafa-DQB1	DQB1*06:08	DQB1*06:08	DQB1*06:08	DQB1*06:08	DQB1*06:08	DQB1*06:08	DQB1*06:08	DQB1*18:07	DQB1*06:08	DQB1*18:07	DQB1*06:08	DQB1*06:08
	Mafa-DPA1	DPA1*02:05	DPA1*02:05	DPA1*02:05	DPA1*02:05	DPA1*02:05	DPA1*02:05	DPA1*02:05	DPA1*04:02	DPA1*02:05	DPA1*02:15:02	DPA1*02:05	DPA1*02:05



c

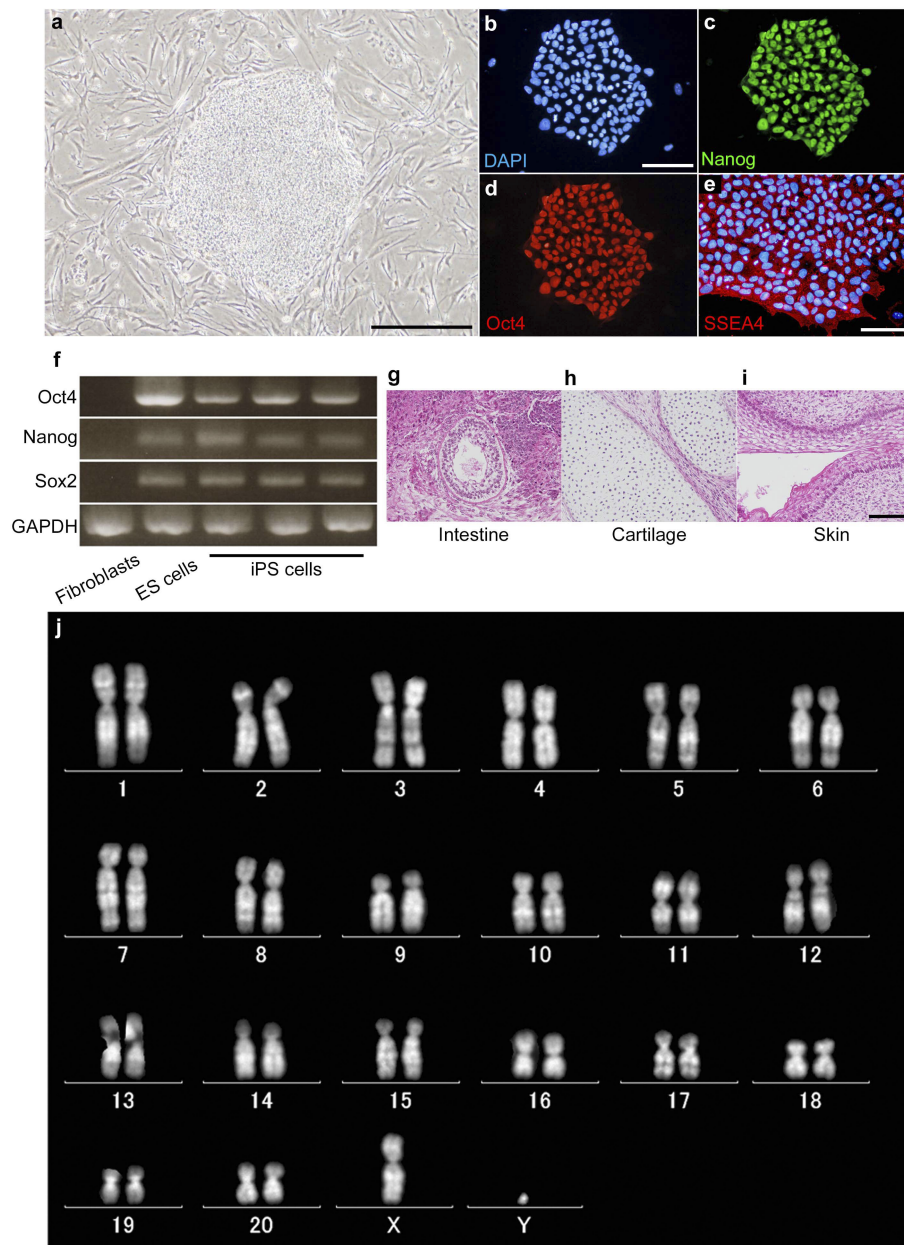


Extended Data Figure 1 | Characteristics of the HT4 haplotype.

a, b, Basic structure of MHC in HT4 haplotypes. One of the cynomolgus monkeys (DrpZ5-32B-C) is strictly a 'homozygote' that has the A-Hp7.2 and B-Hp2 haplotypes in the Mafa-class I region and the #7 haplotype in the Mafa-class II region on both chromosomes (tentatively named 'HT4').

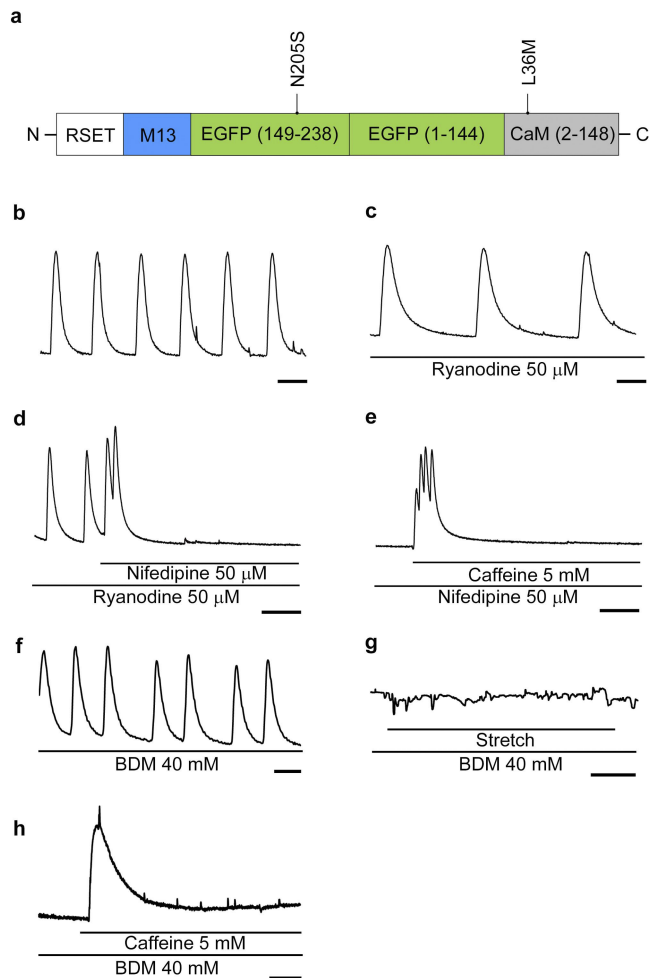
c, *In vitro* mixed lymphoid reactions (MLR) showed that when inactivated

lymphocytes from a HT4-heterozygous monkey were cocultured with active lymphocytes from a HT4-homozygous monkey, proliferation was inhibited to the level of control (only inactivated cells) or autologous (inactivated and active cells from same animal). 'MHC mismatched' indicates two groups of lymphocytes from two different animals with different MHC types. ** $P < 0.01$ versus control. $n = 5$ per group.

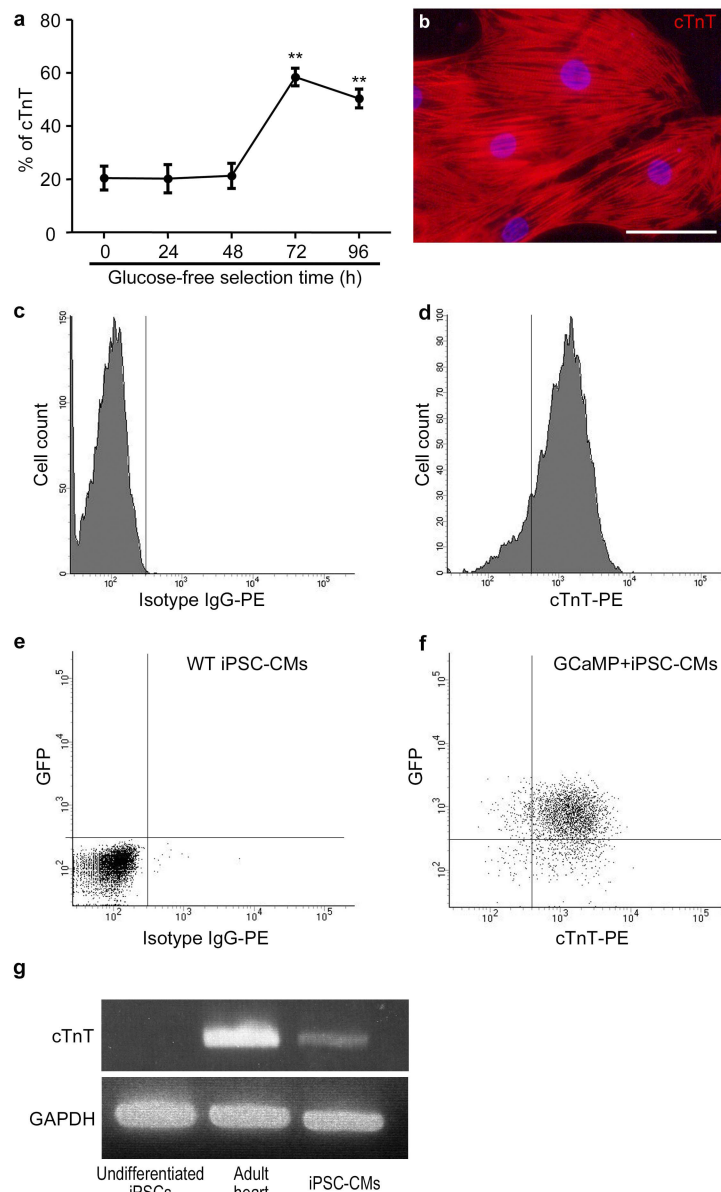


Extended Data Figure 2 | Generation of iPSCs from a MHC homologous cynomolgus monkey. Donor iPSCs were established from skin fibroblasts by transfection of episomal vectors carrying *OCT4*, *KLF4*, *SOX2* and *L-MYC*. **a**, iPSCs form typical ES-cell-like colonies. Scale bar, 50 μ m. **b–e**, iPSCs express pluripotent markers as assessed by immunofluorescence. Scale bars, 100 μ m. **f**, Gene expression of pluripotent markers in the iPSCs

is identical to that in cynomolgus ES cells. **g–i**, When transplanted into immunodeficient mice, the iPSCs gave rise to teratomas manifesting all three germ layers: endoderm (intestinal epithelium), mesoderm (cartilage) and ectoderm (squamous cells). **j**, After expansion, the iPSCs showed normal karyotype (42, XY).

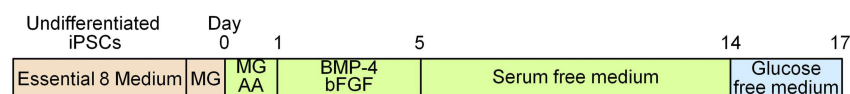
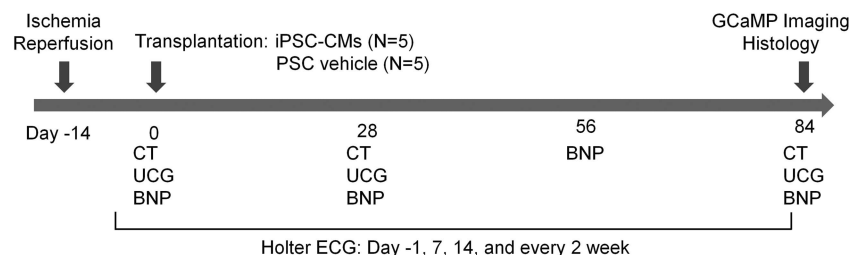


Extended Data Figure 3 | Characteristics of G-CaMP7.09. **a**, Schematic structure of G-CaMP7.09. Mutations are indicated with respect to G-CaMP7. RSET and M13 are tags that encode hexahistidine and a target peptide for Ca^{2+} -bound CaM derived from myosin light chain kinase, respectively. The amino-acid numbers of EGFP and CaM are indicated in parentheses. The dynamic range of G-CaMP7.09 ($F_{\text{max}}/F_{\text{min}}$) was 19.3 ± 2.52 ($n = 3$) and the K_d for Ca^{2+} was 212 ± 6.9 nM ($n = 3$). **b–h**, *In vitro* fluorescence transients of G-CaMP7.09-expressing cardiomyocytes. Data are representative of three independent experiments. **b**, Spontaneous contraction. Scale bar, 2 s. **c**, The firing rate of G-CaMP signals was reduced by treatment with ryanodine, a ryanodine receptor blocker. Scale bar, 2 s. **d**, Addition of the L-type calcium-channel blocker, nifedipine, resulted in cessation of fluorescent transients. Scale bar, 6 s. **e**, Treatment with an activator of the ryanodine receptor, caffeine, induced fluorescent transients in the G-CaMP7.09-expressing iPSC-CMs. Scale bar, 6 s. **f**, G-CaMP7.09 transients were sustained for a few minutes after spontaneous contraction and stopped by 40 mM BDM. Scale bar, 1 s. **g**, After cessation of spontaneous fluorescent transients, iPSC-CMs on Parafilm were stretched but no fluorescent transient was detected. Scale bar, 10 s. **h**, Treatment with caffeine induced G-CaMP7.09 transients again. Scale bar, 5 s.



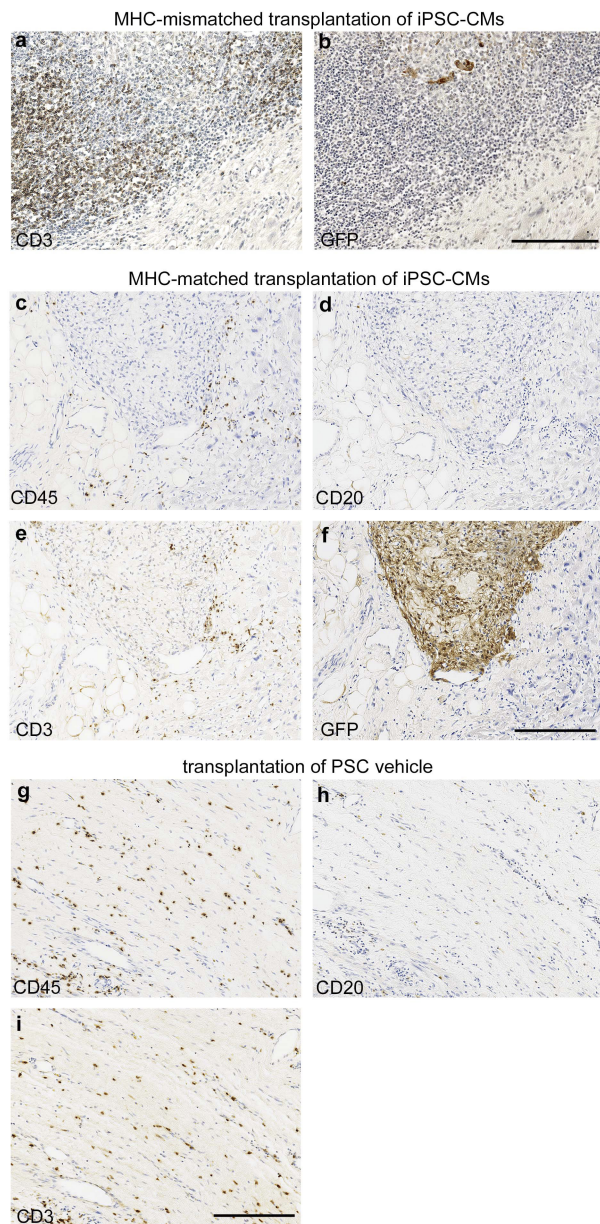
Extended Data Figure 4 | Generation and purification of cynomolgus iPS cell-derived cardiomyocytes. **a**, Pilot experiments showed that cultivation of iPSC-CMs in glucose-free medium for 72 h significantly enhances cardiac purity, $**P < 0.01$ versus 0 h, $n = 4$ for each time point. Data are representative of three independent experiments. **b**, iPSC-CMs express the cardiac-specific marker cTnT. Scale bar, 50 μm .

c, d, After multiple attempts to generate cardiomyocytes for transplantation, 2×10^9 cardiomyocytes (cTnT-positive $83.8 \pm 1.0\%$ as indicated by flow-cytometric analysis) were prepared. **e, f**, The cardiomyocytes were positive for GFP. **g**, RT-PCR analysis indicated that cTnT mRNA expression in iPSC-CMs was detectable, but lower than in the adult heart. Data are representative of three independent experiments.

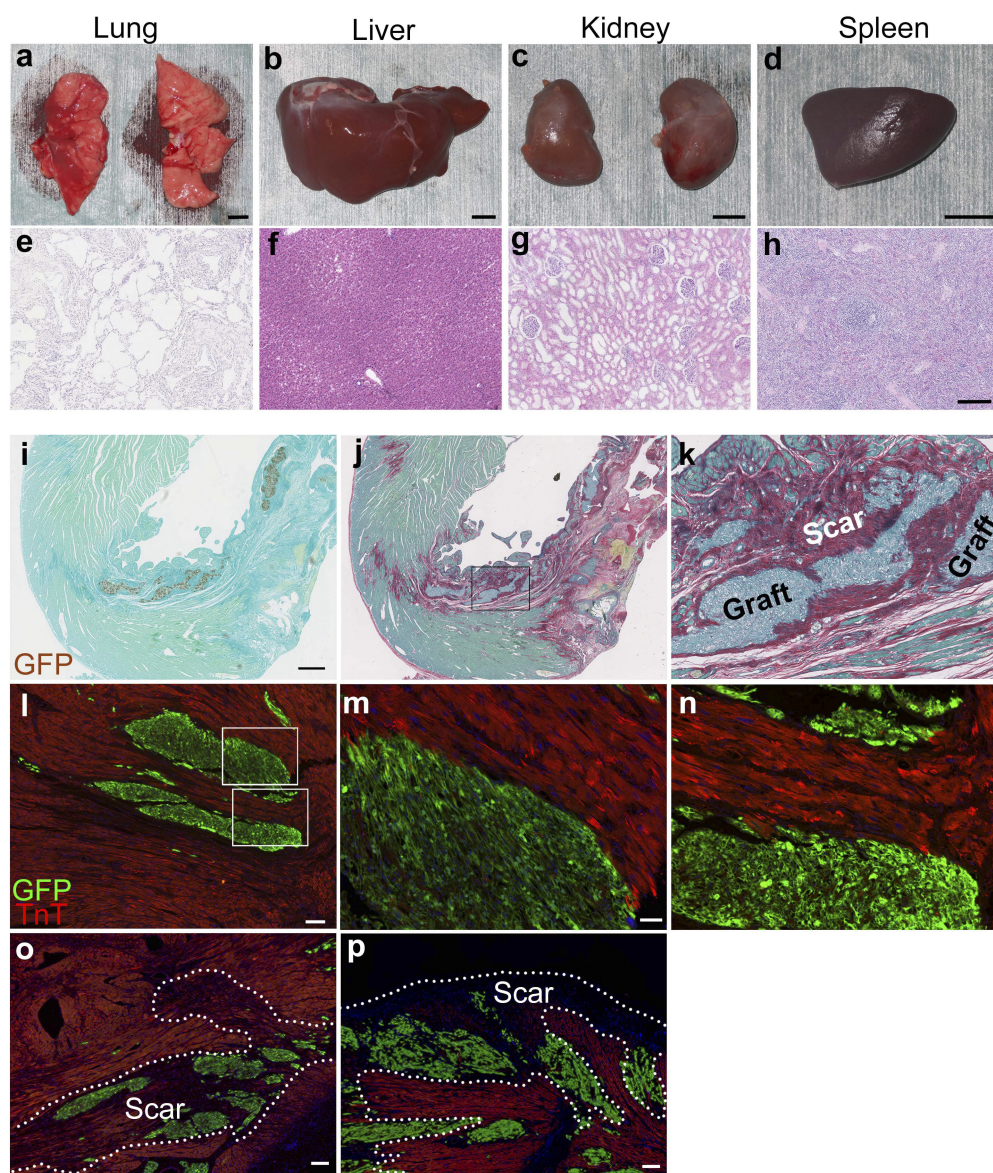
a. Cardiac differentiation protocol**b. In vivo transplantation study protocol**

Extended Data Figure 5 | Study protocol and design. a, A monolayer of cultured undifferentiated cynomolgus monkey iPSCs on a Matrigel (MG)-coated dish was treated with Matrigel. The culture medium was replaced with serum-free medium supplemented with Matrigel and activin A (AA) on day 0. On day 1 after activation, the medium was replaced with medium containing BMP4 and basic fibroblast growth factor (bFGF), and cells were cultured until day 5. On day 14, cardiomyocytes were selected by cultivation in glucose-free medium for 3 days. **b,** Fourteen days before transplantation, 10 female monkeys

were subjected to ischaemia/reperfusion injury. Either 4×10^8 iPSC-CMs reconstituted in a prosurvival cocktail (PSC) or the PSC vehicle was injected on day 0. Cardiac μ CT and UCG were performed to evaluate cardiac contractile function before and after transplantation. Additionally, BNP was measured. Spontaneous arrhythmias were monitored by Holter electrocardiogram (ECG) on days -1, 7, 14 and every other week thereafter. On day 84, all animals were euthanized, and the hearts were excised and subjected to intravital G-CaMP imaging, followed by histological analysis.



Extended Data Figure 6 | Immune response following transplantation of iPS cell-derived cardiomyocytes. **a, b,** iPSC-CMs were transplanted into MHC-mismatched infarcted hearts ($n = 2$). Animals were euthanized and the hearts were collected at 4 weeks post-transplantation. Only a small portion of grafts (GFP) showed a severe infiltration of inflammatory cells, such as CD3⁺ T lymphocytes. **c–i,** Immunohistochemical analysis of recipients of iPSC-CMs or PSC vehicle 84 days post-transplantation. The sections were stained with antibodies against CD45 (leukocytes), CD20 (B lymphocytes), CD3 (T lymphocytes) and GFP (graft). Scale bars in **a–i,** 200 μm .

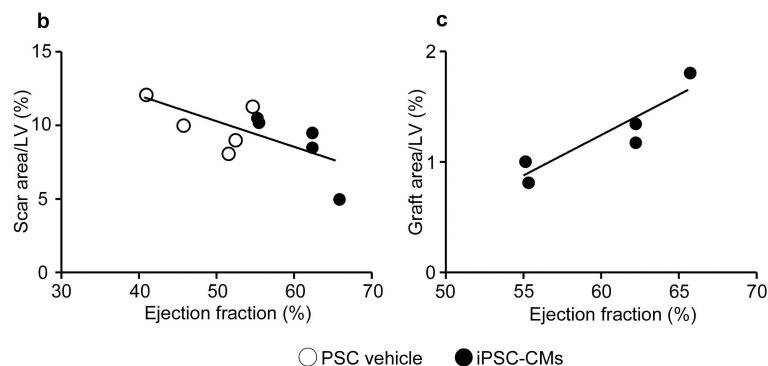


Extended Data Figure 7 | Macroscopic and microscopic analysis of iPSC-CM recipients. **a–h**, All recipients of iPSC-CMs received full necropsy after euthanasia. Neither macroscopic (**a–d**) nor microscopic (**e–h**) analysis revealed any evidence of tumour formation at 12 weeks post cell transplantation. Scale bars in **a–d** and **e–h**: 10 mm and 200 μm, respectively. **i–p**, Additional immunohistochemical analysis of cynomolgus hearts. **i**, Immunohistochemistry for GFP (brown) counterstained with fast green. Scale bar, 1 mm. **j, k**, Picrosirius red staining of a section in

close proximity to the visual field in **a** shows partial remuscularisation of the scar (shown in red) by grafted cardiomyocytes. **l–n**, Different sections (lower by 5 mm towards the apex) showing the corresponding 2 grafts from Fig. 1b. Scale bar, 200 μm. **m, n**, Magnified images of the grafts, scale bar, 50 μm. Note the more direct contact zone of grafted cardiomyocytes with host myocardium. **o, p**, Additional examples of grafted cardiomyocytes in the scar and the border zone. Scale bars, 200 μm.

a

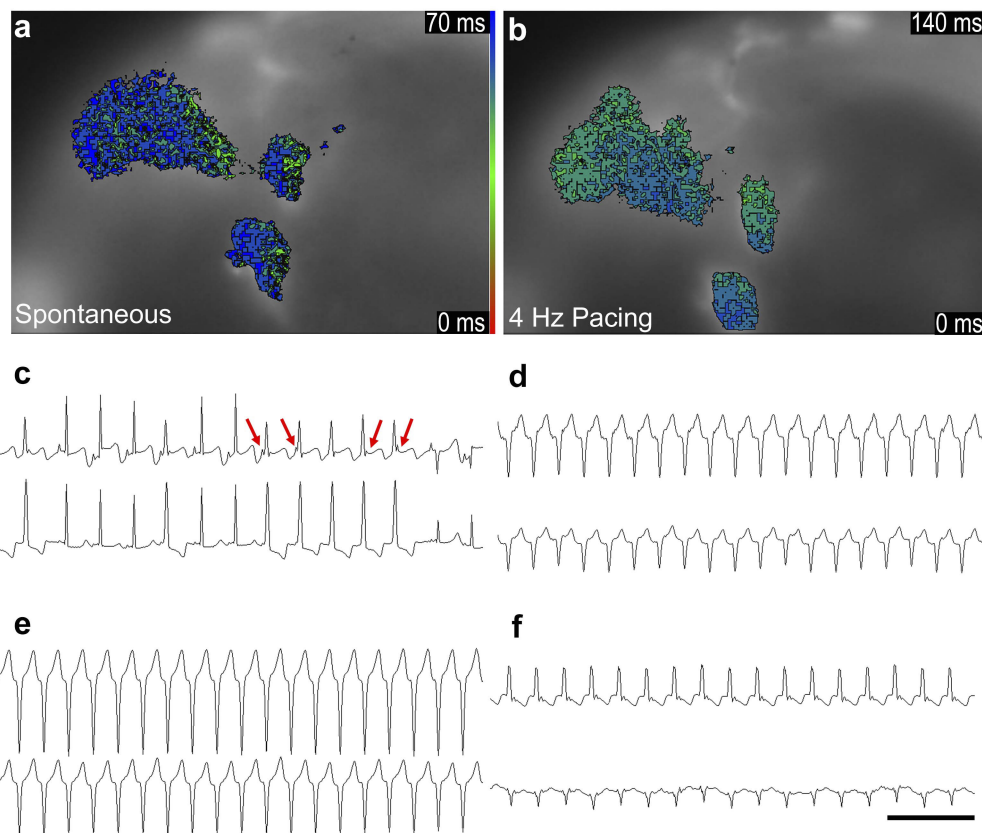
ID	Gender	Age (year)	BW (kg)	Treatment	%Graft area/Scar	%Graft area/LV	%Scar area/LV	EF on Day 84 (%)	# of grafts by histology (scar/total)	# of grafts by intravital imaging (coupled/total)
C034	Female	5	3.2	PSC-vehicle	N/A	N/A	8.2	51.3	N/A	N/A
C035	Female	5	3.45	iPSC-CMs	8.1	0.83	10.3	55.2	3/4	2/2
C036	Female	5	2.83	PSC-vehicle	N/A	N/A	10.1	45.5	N/A	N/A
C037	Female	5	2.93	iPSC-CMs	12.4	1.19	9.6	62.1	4/6	3/3
C038	Female	5	2.6	iPSC-CMs	35.6	1.82	5.1	65.6	8/10	4/4
C039	Female	5	3.01	PSC-vehicle	N/A	N/A	12.2	40.7	N/A	N/A
C040	Female	4	3.14	iPSC-CMs	15.8	1.36	8.6	62.1	5/8	3/3
C041	Female	4	3.21	PSC-vehicle	N/A	N/A	11.4	54.4	N/A	N/A
C042	Female	4	2.88	PSC-vehicle	N/A	N/A	9.1	52.2	N/A	N/A
C043	Female	4	2.87	iPSC-CMs	9.6	1.02	10.6	55	3/4	2/2

**d**

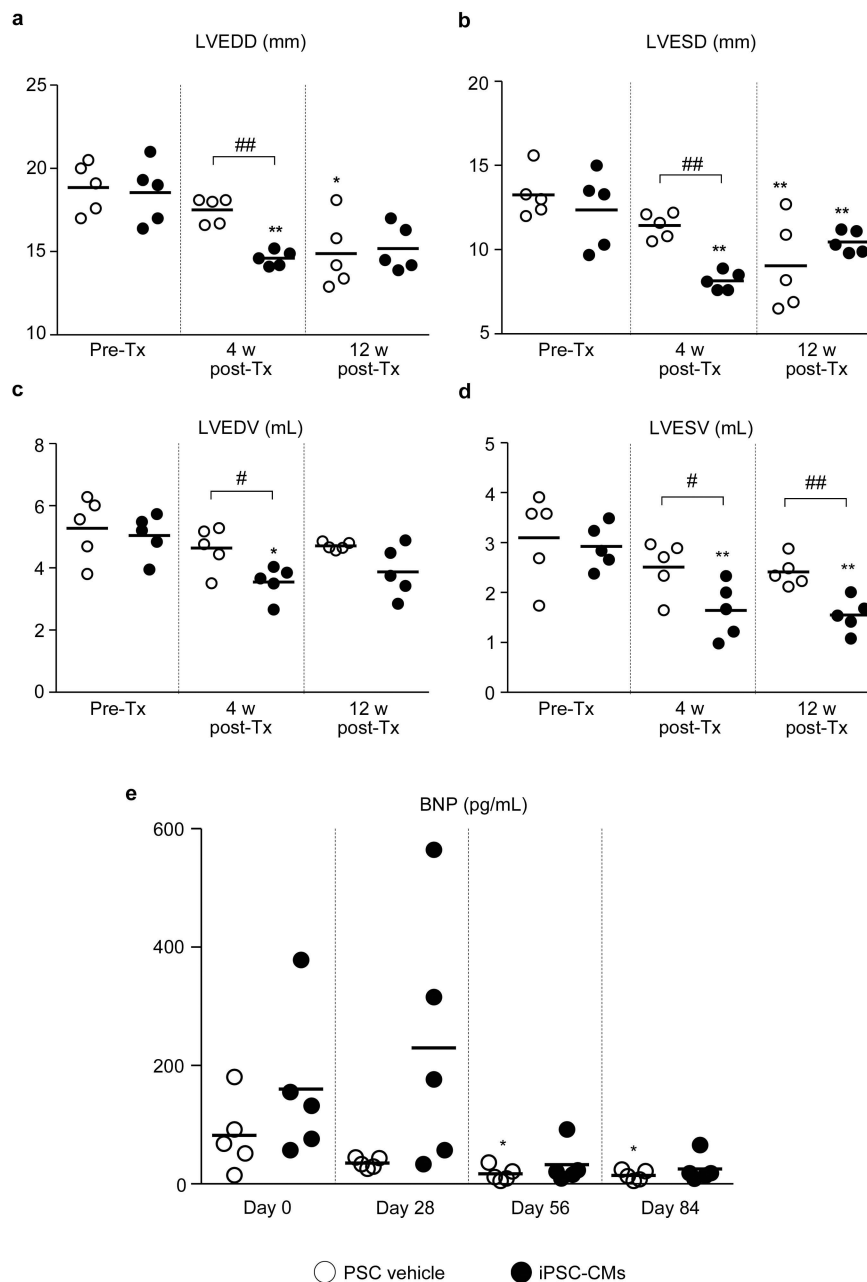
ID	day7			day14			day28			day42			day70		
	# of VTs	max duration (h:m:s)	max HR	# of VTs	max duration (h:m:s)	max HR	# of VTs	max duration (h:m:s)	max HR	# of VTs	max duration (h:m:s)	max HR	# of VTs	max duration (h:m:s)	max HR
C035	208	1:28:29	212	1	24:00:00	249	13	1:56	187	0	N/A	N/A	0	N/A	N/A
C037	149	3:29:54	223	1	24:00:00	249	74	12:52:00	216	91	0:22:25	211	34	0:22:15	216
C038	0	N/A	N/A	3	0:00:58	184	0	N/A	N/A	3	0:00:58	180	0	N/A	N/A
C040	0	N/A	N/A	3	0:00:59	184	83	0:53:45	225	2	0:00:58	180	0	N/A	N/A
C043	0	N/A	N/A	48	1:24:52	202	0	N/A	N/A	0	N/A	N/A	0	N/A	N/A

Extended Data Figure 8 | Summary of histological, mechanical and electrophysiological consequences. **a**, Animal characteristics with histological, mechanical and calcium imaging results. **b**, Correlation between ejection fraction (EF) and scar area relative to left ventricular area

(LV). **c**, Correlation between ejection fraction and graft area relative to left ventricle. **d**, Summary of sustained ventricular tachycardia (VT), including number of VTs, maximum duration, and maximum heart rate (HR), in the recipients of iPSC-CMs.



Extended Data Figure 9 | Additional electrical analysis of hearts transplanted with iPSC-CMs. **a, b,** Activation map obtained from G-CaMP7.09 transients showing the interval (in ms) between the R wave of ECG and the peak of the G-CaMP7.09 fluorescent signal. **c–f,** Examples of sustained and non-sustained VT in recipients of iPSC-CMs. Arrows indicate P wave during VT, suggesting atrioventricular dissociation. Scale bar, 1 s.



Extended Data Figure 10 | Time course of left ventricular size and BNP levels. **a–d**, Left ventricular size was analysed before transplantation (Pre-Tx), 4 weeks post-transplantation (4 w post-Tx) and 12 weeks post-transplantation (12 w post-Tx) by echocardiography (**a**, **b**) and μ CT (**c**, **d**). LVEDD: left ventricular end-diastolic dimension, LVESD: left ventricular end-systolic dimension, LVEDV: left ventricular end-diastolic volume,

LVESV: left ventricular end-systolic volume. $n = 5$ per group. $^{\#}P < 0.05$; $^{##}P < 0.01$. $^{*}P < 0.05$; $^{**}P < 0.01$ versus Pre-Tx. **e**, BNP was measured on days 0 (14 days after myocardial infarction), 28, 56 and 84. No significant difference was detected between recipients of iPSC-CMs and recipients of PSC vehicle at any time point. $^{*}P < 0.05$ versus day 0.

Fetal liver endothelium regulates the seeding of tissue-resident macrophages

Pia Rantakari^{1*}, Norma Jäppinen^{1*}, Emmi Lokka¹, Elias Makkala¹, Heidi Gerke¹, Emilia Peuhu², Johanna Ivaska^{2,3}, Kati Elima^{1,4}, Kaisa Auvinen¹ & Marko Salmi^{1,5}

Macrophages are required for normal embryogenesis, tissue homeostasis and immunity against microorganisms and tumours^{1–4}. Adult tissue-resident macrophages largely originate from long-lived, self-renewing embryonic precursors and not from haematopoietic stem-cell activity in the bone marrow^{4,5}. Although fate-mapping studies have uncovered a great amount of detail on the origin and kinetics of fetal macrophage development in the yolk sac and liver^{6–11}, the molecules that govern the tissue-specific migration of these cells remain completely unknown. Here we show that an endothelium-specific molecule, plasmalemma vesicle-associated protein (PLVAP), regulates the seeding of fetal monocyte-derived macrophages to tissues in mice. We found that PLVAP-deficient mice have completely normal levels of both yolk-sac- and bone-marrow-derived macrophages, but that fetal liver monocyte-derived macrophage populations were practically missing from tissues. Adult PLVAP-deficient mice show major alterations in macrophage-dependent iron recycling and mammary branching morphogenesis. PLVAP forms diaphragms in the fenestrae of liver sinusoidal endothelium during embryogenesis, interacts with chemoattractants and adhesion molecules and regulates the egress of fetal liver monocytes to the systemic vasculature. Thus, PLVAP selectively controls the exit of macrophage precursors from the fetal liver and, to our knowledge, is the first molecule identified in any organ as regulating the migratory events during embryonic macrophage ontogeny.

Tissue-resident macrophages in adults are largely generated in sequential waves during embryogenesis^{1–13}. The first erythro-myeloid progenitors (EMPs) are found in the blood islands of the extra-embryonic yolk sac at embryonic day 7.0 (E7.0)¹⁴. These early EMPs differentiate to yolk-sac macrophages and, after the establishment of blood circulation at E8.5 (ref. 15), migrate to all tissues, including the central nervous system and liver^{6–8,16}. In parallel, Myb-dependent EMPs, mainly generated in the haemogenic endothelium of the yolk sac, seed the fetal liver at E9.5 to generate myeloid progenitors, which give rise to the first fetal monocytes around E12.5 (refs 10, 17). The fetal liver-derived monocytes enter the blood, infiltrate all tissues (except the central nervous system, which has already been isolated by the blood–brain barrier) and differentiate into tissue-resident macrophages, superseding the yolk-sac-derived macrophages after E16.5 (refs 8, 10, 11, 18–21). Simultaneously, haematopoietic stem cells (HSCs) are emerging at sites of intra-embryonic haemogenic endothelium, including at the aorta–gonad–mesonephros (AGM), seeding the fetal liver after E11.5 and contributing to the generation of macrophages via monocytic intermediates^{9,10,17,22}. Around the time of birth and during postnatal life, bone marrow HSC-derived monocytes can give rise to tissue-resident macrophages in selected organs^{19,23}.

In the course of studying the immunological functions of the endothelial-specific molecule PLVAP²⁴, we found major unexpected

alterations in the macrophage system. In adult PLVAP-deficient mice, the frequencies of embryonic-derived, tissue-resident macrophages (CD11b⁺F4/80^{high} cells^{7,11,20}; full flow cytometry gates are shown in Extended Data Fig. 1 and Supplementary Table 1) were reduced by 70% in the spleen and by 95% in the peritoneal cavity when compared to sex-matched, wild-type littermate controls (Fig. 1a). Similarly, the embryonic-derived alveolar macrophage population (CD11b⁺CD11c^{high} cells^{7,11,18,20}) in the adult lungs was significantly diminished in *Plvap*^{−/−} mice (Fig. 1a). By contrast, the frequency of adult bone-marrow-derived tissue macrophages (CD11b⁺F4/80^{intermediate} cells^{7,11,20,23}), which partially (in the spleen, peritoneal cavity and peripheral lymph nodes^{11,20,25}) or completely (in the colon²³) replace embryonically derived macrophages, was normal or saw a compensatory increase in adult *Plvap*^{−/−} mice (Fig. 1a and Extended Data Fig. 2a). Moreover, the frequency of HSCs (Lin[−]c-Kit⁺Sca-1⁺ cells), common myeloid (Lin[−]c-Kit⁺Sca-1[−]IL7R[−]) and common lymphoid (Lin[−]c-Kit⁺Sca-1^{low}IL7R⁺) progenitor cells remained unchanged in the bone marrow of *Plvap*^{−/−} mice; so too did the colony-forming capacity of the bone marrow cells, and the frequencies of monocytes in the bone marrow and blood (both the patrolling CD11b⁺Ly6C^{low} and the tissue-infiltrating CD11b⁺Ly6C^{high} subpopulations²⁶; Fig. 1a and Extended Data Fig. 2b–d). The frequency of recently entered tissue monocytes (CD11b⁺F4/80^{intermediate}Ly6C^{low} and CD11b⁺F4/80^{intermediate}Ly6C^{high} cells^{11,20}) in the spleen and liver was also unchanged in mutant compared to wild-type mice (Extended Data Fig. 2e, f). The frequency of the major lymphocyte subpopulations in adult *Plvap*^{−/−} mice, and of embryonic-derived macrophages in three other gene-deficient animals (*Nt5e*^{−/−} and *Aoc3*^{−/−} with perturbed leukocyte trafficking, and caveolae-deficient *Cav1*^{−/−}), were similar to those in wild-type controls (Extended Data Figs 2g, 3a, b). Thus, PLVAP selectively controls the accumulation of embryonically derived macrophages but is dispensable for the production of adult bone-marrow-derived macrophages and other leukocyte types.

We noticed that the frequency of yolk-sac-derived splenic and lung macrophages (CD11b⁺F4/80^{high} cells^{7,11,20}) was unaffected in *Plvap*^{−/−} embryos, whereas that of fetal liver monocyte-derived macrophages (CD11b⁺Ly6C⁺F4/80^{intermediate} cells^{7,10,11,20}) was clearly lower at E16.5 (Fig. 1b). The frequency of liver-derived monocytes (CD11b⁺Ly6C^{high} cells²⁶) in the blood at E16.5 was also reduced by 55% in *Plvap*^{−/−} mice (Fig. 1b). These data suggest that PLVAP regulates liver-derived, rather than monocyte-independent yolk-sac-derived, macrophage generation.

In more detailed analyses of fetal yolk-sac-, AGM- and liver-derived macrophage production, we observed PLVAP expression in the endothelial cells of E8.5 yolk-sac blood islands (as well as in other endothelial beds of E9.5 fetuses; Extended Data Fig. 4a and Supplementary Video 1). However, the yolk-sac production of early E10.5 and late E12.5 EMPs (CD41⁺CD45⁺c-Kit^{high}F4/80[−] cells^{10,17}, which did not express *Plvap*, and the colony-forming potential of

¹MediCity Research Laboratory, University of Turku, Turku, FI-20520, Finland. ²Turku Centre for Biotechnology, University of Turku, Turku, FI-20520, Finland. ³Department of Biochemistry, University of Turku, Turku, FI-20500, Finland. ⁴Department of Medical Biochemistry and Genetics, University of Turku, Turku, FI-20520, Finland. ⁵Department of Medical Microbiology and Immunology, University of Turku, Turku, FI-20520, Finland.

*These authors contributed equally to this work.

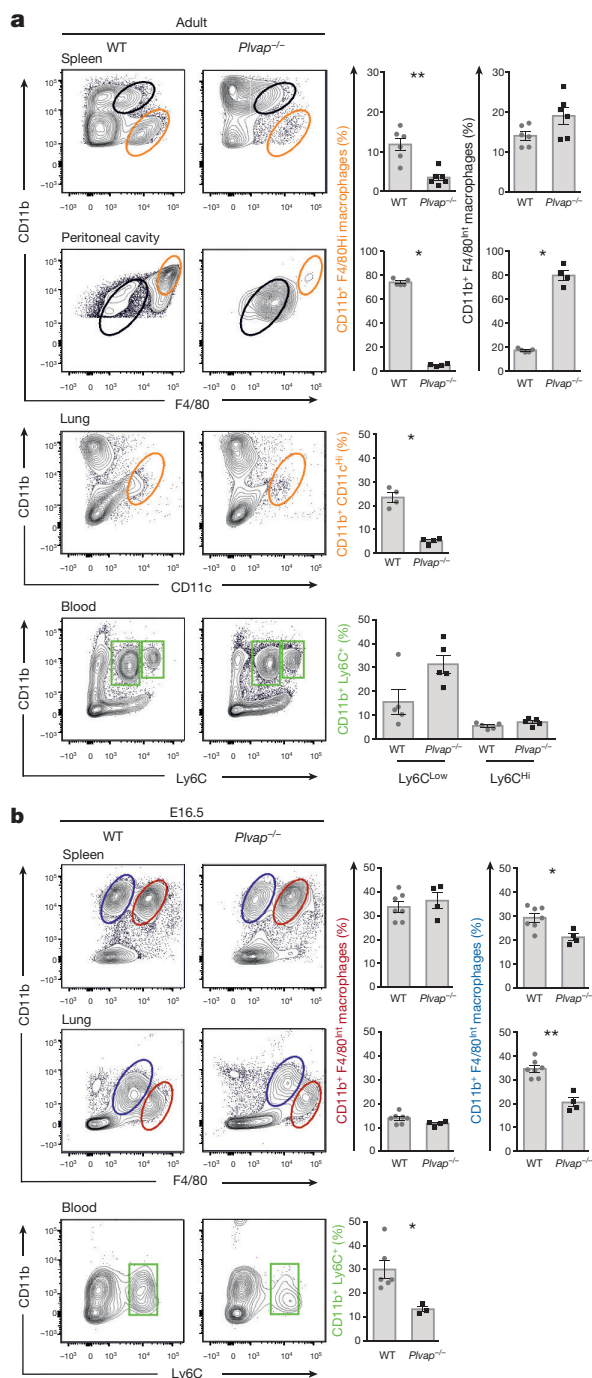


Figure 1 | PLVAP regulates the accumulation of embryonic-derived tissue-resident macrophages. a, b, Flow cytometry analyses of tissue-resident macrophages in the spleen, peritoneal cavity and lung, and of blood monocytes, in adult (a) and E16.5 (b) wild-type (WT) and *Plvap*^{-/-} mice. Adults: orange gates, embryonic-derived macrophages (CD11b⁺F4/80^{high} cells in the spleen and peritoneal cavity and CD11b⁺CD11c^{high} cells in the lungs); black gates, adult bone-marrow-derived CD11b⁺F4/80^{intermediate} macrophages. Embryos: red gates, yolk-sac-derived CD11b⁺F4/80^{high} macrophages; blue gates, fetal liver monocyte-derived CD11b⁺F4/80^{intermediate} macrophages. Blood: green gates, CD11b⁺Ly6C^{high} inflammatory and CD11b⁺Ly6C^{low} patrolling monocytes. The flow cytometry data are shown as the frequency of live-gated CD45⁺B220⁻CD4⁻CD8⁻ (a, adult tissues), and of live-gated CD45⁺B220⁻ (a, adult blood and b) cells. Each dot represents one mouse or embryo (pooled from 2–3 independent experiments and from 4 litters (b), see Source Data), data are mean \pm s.e.m. for each group (* P < 0.05, ** P < 0.01 by Mann–Whitney U -test).

E10.5 yolk sac cells, was comparable in *Plvap*^{-/-} and wild-type mice (Extended Data Fig. 4b–e). Similarly, F4/80⁺ macrophage populations in the yolk sac at E10.5 and E12.5 and their direct descendants, E16.5 brain microglia and adult brain microglia (CD45^{low}F4/80^{high} cells^{6,7,20}), were indistinguishable in *Plvap*^{-/-} and control mice (Fig. 2a and Extended Data Fig. 4b, c). PLVAP was also expressed in the AGM endothelium at E10.5, whereas c-Kit⁺ cells in AGM did not express PLVAP and were not affected by the absence of PLVAP (Extended Data Fig. 4f, g). Thus, the emergence of primitive, yolk-sac-derived macrophage progenitors and macrophages, as well as early c-Kit⁺ cells in the AGM, is a PLVAP-independent process.

Following yolk-sac production, embryonic macrophage generation shifts to the liver from E12.5 (refs 10, 17). The frequency of yolk-sac-derived macrophages (CD11b⁺F4/80^{high} cells^{7,10,11,20}) in E12.5, E14.5 and E16.5 liver was similar in PLVAP-deficient and wild-type mice (Fig. 2b). By contrast, after E14.5 there were more liver-derived macrophages (CD11b⁺Ly6C⁺F4/80^{intermediate} cells^{7,10,11,20}) found in *Plvap*^{-/-} livers than in those of wild-type controls (Fig. 2b). When sorted from fetal livers, yolk-sac-derived CD11b⁺F4/80^{high} macrophages and fetal liver monocyte-derived CD11b⁺F4/80^{intermediate} cells showed the previously reported differences in morphology and gene expression signatures^{7,8} (Extended Data Fig. 5a, b). In closer analyses, the frequency of fetal liver monocytes (both CD11b⁺CSF-1R⁺c-Kit⁺Flt-3⁻Ly6C⁺ and CD11b⁺CSF-1R⁺c-Kit⁻Flt-3⁻Ly6C⁻ monocytes¹⁰) was significantly increased (by 35–75%) at E13.5–E16.5 in *Plvap*^{-/-} mice (Fig. 2c, d and Extended Data Fig. 5c). However, the frequencies of all monocyte progenitor cell types (that is, CD11b⁺CSF-1R⁺c-Kit⁺Flt-3⁺Ly6C⁻ macrophage–dendritic cell precursors, CD11b⁺CSF-1R⁺c-Kit⁺Flt-3⁻Ly6C⁻ fetal liver myeloid precursors, and CD11b⁺CSF-1R⁺c-Kit⁺Flt-3⁻Ly6C⁺ common monocyte progenitors¹⁰) were comparable between the *Plvap*^{-/-} and wild-type livers at E12.5–E16.5 (Extended Data Fig. 5c). In *Plvap*^{-/-} livers, neither the frequency of EMPs (CD41⁺CD45⁺c-Kit^{high}F4/80⁻ cells^{10,17}) and HSCs (Lin⁻c-Kit⁺Sca-1⁺ cells⁹) nor their colony-forming ability was altered at E12.5 (Extended Data Fig. 5d–f). Notably, prevention of yolk-sac macrophage (but not EMP) development by a single injection of anti-CSF-1R antibody (clone AFS) to E6.5 pregnant mice¹⁰ did not blunt the increased accumulation of Ly6C⁻ and Ly6C⁺ monocytes in the fetal *Plvap*^{-/-} liver (Extended Data Fig. 6a–c). In the adult liver, the frequency of embryonic-derived macrophages (CD11b⁺F4/80^{high} cells^{10,11,20}) was lower, whereas that of bone-marrow-derived CD11b⁺F4/80^{intermediate} macrophages was higher in *Plvap*^{-/-} than in control mice (Fig. 2b). These data, together with the low blood monocyte counts at E16.5 (Fig. 1b), suggest that the entry and differentiation of macrophage precursors in fetal liver is intact in *Plvap*^{-/-} mice, whereas the exit of mature fetal liver monocytes is impaired in the absence of PLVAP.

PLVAP protein and *Plvap* mRNA were synthesized in the wild-type liver from E11.5 onwards, and were completely absent from *Plvap*^{-/-} livers (Fig. 3a, b, Extended Data Fig. 7a, b and Supplementary Video 2). The overall liver morphology and vasculature were indistinguishable between *Plvap*^{-/-} and wild-type mice (Extended Data Fig. 7a–c). Notably, during embryogenesis PLVAP was only expressed in LYVE-1⁺ and CD31⁺ endothelial cells, and was completely absent from hepatocytes, leukocytes (including the F4/80⁺ cells, which are in close contact with the endothelium) and other stromal cells in the liver (Fig. 3a and Extended Data Fig. 7a, d–g). PLVAP protein was found on the luminal surface of liver endothelial cells at E12.5, since an intravascularly administered anti-PLVAP antibody MECA-32, but not an isotype-matched negative-control antibody, brightly stained the sinusoidal LYVE-1⁺ endothelial cells (Fig. 3c). Selective PLVAP expression in fetal liver sinusoidal endothelial cells was also detected in other mouse strains and in humans (Extended Data Fig. 7h, i).

In the blood vascular endothelial cells, PLVAP is only present in and is the sole component of diaphragms (cartwheel-like flat structures), which can overlay fenestrations, transendothelial channels

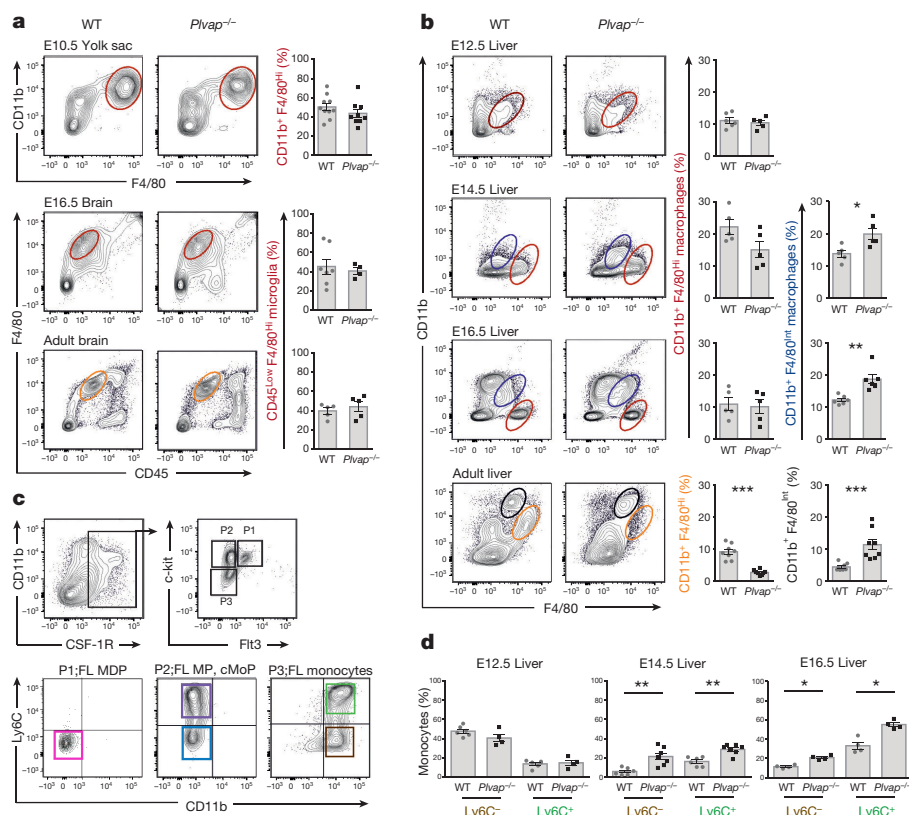


Figure 2 | Fetal liver-dependent, but not yolk-sac-dependent, macrophage seeding is controlled by PLVAP. **a**, Flow cytometry analyses of CD11b⁺F4/80^{high} yolk-sac macrophages at E10.5, and of CD45^{low}F4/80^{high} microglia in E16.5 (red gates) and adult (orange gates) brain in wild-type and *Plvap*^{-/-} mice. **b**, Flow cytometry analyses of yolk-sac-derived CD11b⁺F4/80^{high} (red gates), and fetal liver-derived CD11b⁺F4/80^{intermediate} (blue gates) macrophages in fetal livers, and of embryonic-derived CD11b⁺F4/80^{high} (orange gates) and adult bone-marrow-derived CD11b⁺F4/80^{intermediate} (black gates) macrophages in adult livers. **c**, **d**, The gating strategy to identify fetal liver CD11b⁺CSF-1R⁺c-Kit⁻Flt-3⁻Ly6C⁻ monocytes (Ly6C⁻ monocytes, first P3, then the brown gate) and CD11b⁺CSF-1R⁺c-Kit⁻Flt-3⁻Ly6C⁺ monocytes (Ly6C⁺ monocytes, first P3, then the green gate) (**c**), and their enumeration (**d**). cMoP, common monocyte progenitor, FL, fetal liver; MDP, macrophage dendritic cell precursor, MP, myeloid precursor. For the full gating strategies, see Extended Data Figs. 1 and 5c and Supplementary Table 1. Each dot represents one mouse or embryo (pooled from 2–4 independent experiments and from 2–5 litters (**a–d**), see Source Data), data are mean \pm s.e.m. for each group (* P < 0.05, ** P < 0.01, *** P < 0.001 by Mann–Whitney U -test).

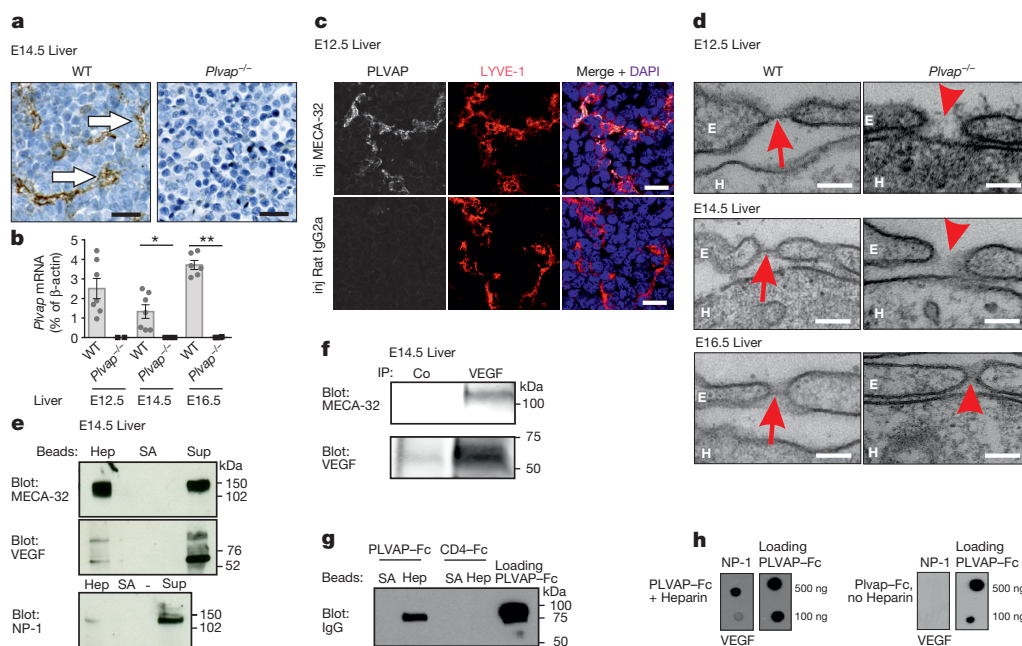


Figure 3 | PLVAP forms diaphragms in the fetal liver sinusoidal endothelial cells and interacts with heparin, VEGF and neuropilin-1. **a**, **b**, PLVAP protein (**a**, MECA-32 immunostain) and mRNA (**b**) expression in fetal livers. White arrows, representative vessels. **c**, Detection of luminal PLVAP on liver sinusoidal endothelial cells after intravascular injections of MECA-32 or control (rat IgG2a) antibody to wild-type mice. The sections were stained *ex vivo* for LYVE-1. **d**, Transmission electron micrographs of fetal liver sinusoidal endothelium. Red arrows, diaphragms; red arrowheads, fenestrae without diaphragms; E, endothelial cell; H, hepatocyte. **e**, Pull-down assays with heparin-affinity (Hep) and streptavidin-affinity (SA, a negative control) beads from lysates of wild-type E14.5 livers for PLVAP (MECA-32), VEGF and neuropilin-1 (NP-1). Sup, liver lysate (loading control). **f**, Co-immunoprecipitation assays from

E14.5 liver. IP, immunoprecipitation; Co, negative control antibody. **g**, Pull-down assays with heparin and streptavidin beads for PLVAP-Fc and CD4-Fc (a negative control) fusion proteins. **h**, Far-western assays for binding of PLVAP-Fc to recombinant neuropilin-1 (250 ng per spot) and VEGF (250 ng per spot) in the presence and absence of heparin (50 IU). Loading PLVAP-Fc denotes loading controls. Shown are representative images ($n = 2$ and 3 biological replicates for **a** and **d–h**, respectively; $n = 4$ (MECA-32) and $n = 1$ (rat IgG2a) biological replicates (**c**)). Scale bars, 20 μ m (**a**, **c**), 100 nm (**d**). Each dot represents one embryo (pooled from 2–3 independent experiments and from 2–4 litters (**b**); see Source Data), data are mean \pm s.e.m. of each group (* P < 0.05 by Mann–Whitney U -test). The gel source data are provided in Supplementary Fig. 1.

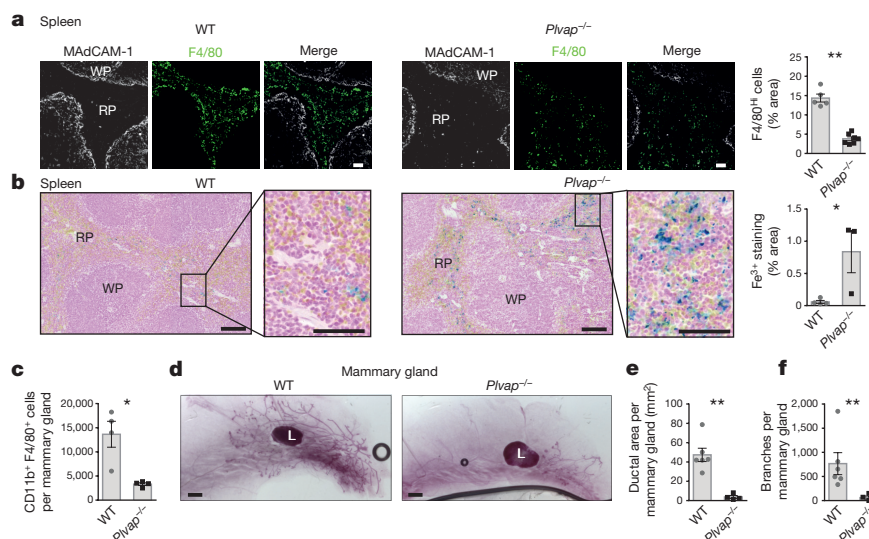


Figure 4 | Altered iron homeostasis and mammary branching morphogenesis in *Plvap*^{-/-} mice. **a**, Immunofluorescent stains of adult spleens for F4/80 (red pulp macrophages) and MAdCAM-1 (the marginal sinus), and quantification of the F4/80^{high} cells. **b**, Prussian blue stains of spleens, and quantification of Fe³⁺-containing cells (blue) in 5-week-old mice. **c–f**, Assessment of mammary gland development in 4.5-week-old mice by flow cytometry analysis of live-gated CD45⁺B220⁻CD11b⁺F4/80⁺ cells (**c**), whole-mount carmine alum stains

of mammary fat pads (**d**), and quantification of mammary gland ductal area (**e**) and ductal branching (**f**). L, lymph node; RP, red pulp; WP, white pulp. Shown are representative images (**a**, **b** (left), **d**; *n* = 4 (**a**), 3 (**b**) and 4–6 (**d**) biological replicates from 2–6 independent stains) and quantitative data (**a**, **b** (right), **c**, **e**, **f**). Scale bars, 50 μ m (**a**), 100 μ m (**b**, main image), 50 μ m (**b**, inset), 1 mm (**d**). Each dot represents one mouse (pooled from 2–4 independent experiments, see Source Data), data are mean \pm s.e.m. for each group (**P* < 0.05, ***P* < 0.01 by Mann–Whitney *U*-test).

and caveolae^{27,28}. Transmission electron microscopy revealed that in wild-type mice the diaphragms distended the fenestrae of liver sinusoidal endothelial cells at E12.5, E14.5 and E16.5 (Fig. 3d). *Plvap*^{-/-} mice showed a complete absence of diaphragms in the liver fenestrae throughout embryogenesis (Fig. 3d). PLVAP also formed diaphragms in certain caveolae in wild-type fetal liver sinusoidal endothelial cells, but the lack of all caveolae in *Cav1*^{-/-} mice affected neither overall PLVAP staining in the liver endothelial cells nor macrophage or monocyte frequencies at E16.5 or in the adults (Extended Data Fig. 3b–e). Thus, the PLVAP-dependent formation of endothelial diaphragms in the fenestrae of sinusoidal endothelium, rather than in caveolae, correlates to the egress of monocytes from the liver.

To dissect the liver-selective role of PLVAP in macrophage ontogeny further we generated a tamoxifen-inducible *Plvap*^{F/F};CAGGCre-ERTM mouse line to delete *Plvap* only after the period of yolk-sac-dependent macrophage production. A single tamoxifen pulse at E12.5 caused a partial reduction in PLVAP protein and mRNA at E14.5 and correlated to a small, yet significant and selective, accumulation of Ly6C⁺ monocytes in the fetal liver, whereas *Plvap* deletion at an earlier (E11.5–E13.5) or later (E13.5–E15.5) time point did not (Extended Data Fig. 8a–d). Analyses of conditional *Plvap*^{F/F};Lyve1-Cre mice, targeting within the blood vasculature only selected vessels, including those in the yolk sac and fetal liver, also revealed a small but selective increase in fetal liver monocytes at E13.5 and E14.5 (Extended Data Fig. 8e, f). Collectively, these data suggest that PLVAP function at E13.5–E14.5 is needed for normal egress of fetal liver monocytes.

Given that sinusoidal endothelial cells with PLVAP diaphragms supported the seeding of fetal liver monocytes much more efficiently than those with open fenestrae, we reasoned that PLVAP could interact with molecules that potentially regulate monocyte emigration. We observed that endogenous PLVAP in E14.5 fetal liver lysates bound to heparin-affinity beads, but not to control beads, in pull-down assays (Fig. 3e). Several chemotactic molecules known to mediate monocyte and macrophage migration in adults, such as VEGF-A²⁹, interact with heparin, raising the possibility that they could interact with PLVAP via a heparin bridge. Endogenous VEGF-A from E14.5 liver selectively bound to heparin-affinity beads, co-immunoprecipitated PLVAP, and

co-localized and closely associated with PLVAP in sinusoidal endothelial cells *in vivo* (Fig. 3e, f and Extended Data Fig. 9a, b). A PLVAP–IgG Fc domain fusion protein (PLVAP–Fc) showed direct, specific and avid interactions with heparin (Fig. 3g and Extended Data Fig. 9c–e). Moreover, it bound to recombinant VEGF in the presence, but not the absence, of heparin in far-western assays (Fig. 3h). In search of VEGF receptors, we observed that in fetal livers, neuropilin-1 is induced on common monocyte progenitors and Ly6C⁻ and Ly6C⁺ monocytes, which also expressed VEGFR1 but not VEGFR2 (Extended Data Fig. 9f–i). Neuropilin-1 is known to be a chemotactic, heparin-binding molecule³⁰. Both recombinant neuropilin-1 and neuropilin-1 expressed by E14.5 liver lysates bound to heparin and interacted with the PLVAP–Fc fusion protein in the presence, but not the absence, of heparin (Fig. 3e, h). Thus, PLVAP–heparin complexes in the fenestral diaphragms may assist emigration of fetal liver monocytes by immobilizing chemotactic molecules (for example, VEGF, for which fetal liver monocytes express VEGFR1 and neuropilin-1 receptors) and/or by providing a substrate for monocyte adhesion molecules (for example, neuropilin-1).

Finally, we studied the functions of macrophages in *Plvap*^{-/-} mice. Macrophage-dependent morphogenetic events that occur during embryogenesis, including bronchial branching and interdigital regression^{1,2,4}, were normal in *Plvap*^{-/-} mice (Extended Data Fig. 10a, b), which is in line with an intact yolk-sac-derived macrophage system. By contrast, we found a reduced number of F4/80^{high} macrophages, which normally mediate the recycling of iron^{2,4,25}, in the red pulp of the spleen (Fig. 4a) and in the liver (Fig. 2b) of 5-week-old *Plvap*^{-/-} mice. Staining with Prussian blue showed increased accumulation of Fe³⁺ in the red pulp of spleen and in the liver in *Plvap*^{-/-} mice (Fig. 4b and Extended Data Fig. 10c). In the mammary fat pads, the number of CD11b⁺F4/80⁺ macrophages was significantly reduced in *Plvap*^{-/-} mice, while B and T lymphocytes were not affected (Fig. 4c and Extended Data Fig. 10d, e). The ductal branching morphogenesis in the mammary glands of prepubertal 4.5-week-old *Plvap*^{-/-} mice failed almost completely (Fig. 4d–f).

Here we have shown that *Plvap*^{-/-} mice have a lower frequency (and total cell number; Supplementary Table 3) of fetal liver monocyte-derived tissue-resident macrophages. By contrast, the frequencies of yolk-sac-derived macrophages, EMPs, HSCs, and bone-marrow-derived

progenitor cells, and bone-marrow-derived monocytes and macrophages are not affected, but these cell types are unable to compensate for the altered macrophage functions in post-natal *Plvap*^{-/-} mice. Notably, during haematopoiesis both the endothelial cells (fenestrated with diaphragms) and monocytes (global gene expression profiles and CCR2-independent egress mechanisms) in the fetal liver have unique characteristics when compared to those in adult bone marrow (Fig. 3d and refs 7, 10). Our results suggest that PLVAP selectively controls the seeding of monocytes at the liver sinusoidal endothelium (although formal proof of this would require the discovery of a liver-sinusoidal-endothelium-specific Cre-mouse), and that PLVAP, previously thought to function solely as a physical filter^{27,28}, has the potential to participate in multiple molecular interactions with heparin-binding chemoattractants and other molecules (Extended Data Fig. 10f). Identification of PLVAP as the first molecule regulating the organ-selective seeding of fetal macrophages may assist in understanding the role of other molecules and the detailed mechanisms involved in these critical migratory steps during embryonic macrophage production.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 March; accepted 31 August 2016.

Published online 12 October 2016.

- Wynn, T. A., Chawla, A. & Pollard, J. W. Macrophage biology in development, homeostasis and disease. *Nature* **496**, 445–455 (2013).
- Davies, L. C., Jenkins, S. J., Allen, J. E. & Taylor, P. R. Tissue-resident macrophages. *Nat. Immunol.* **14**, 986–995 (2013).
- Ginhoux, F., Schultze, J. L., Murray, P. J., Ochando, J. & Biswas, S. K. New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nat. Immunol.* **17**, 34–40 (2016).
- Varol, C., Mildner, A. & Jung, S. Macrophages: development and tissue specialization. *Annu. Rev. Immunol.* **33**, 643–675 (2015).
- Sieweke, M. H. & Allen, J. E. Beyond stem cells: self-renewal of differentiated macrophages. *Science* **342**, 1242–1247 (2013).
- Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
- Schulz, C. *et al.* A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* **336**, 86–90 (2012).
- Gomez Perdiguero, E. *et al.* Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547–551 (2015).
- Sheng, J., Ruedl, C. & Karjalainen, K. Most tissue-resident macrophages except microglia are derived from fetal hematopoietic stem cells. *Immunity* **43**, 382–393 (2015).
- Hoeffel, G. *et al.* C-Myb⁺ erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. *Immunity* **42**, 665–678 (2015).
- Yona, S. *et al.* Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* **38**, 79–91 (2013).
- Okabe, Y. & Medzhitov, R. Tissue biology perspective on macrophages. *Nat. Immunol.* **17**, 9–17 (2016).
- Amit, I., Winter, D. R. & Jung, S. The role of the local environment and epigenetics in shaping macrophage identity and their effect on tissue homeostasis. *Nat. Immunol.* **17**, 18–25 (2016).
- Palis, J., Robertson, S., Kennedy, M., Wall, C. & Keller, G. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* **126**, 5073–5084 (1999).
- McGrath, K. E., Koniski, A. D., Malik, J. & Palis, J. Circulation is established in a stepwise pattern in the mammalian embryo. *Blood* **101**, 1669–1676 (2003).
- Kierdorf, K. *et al.* Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16**, 273–280 (2013).
- McGrath, K. E. *et al.* Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. *Cell Reports* **11**, 1892–1904 (2015).
- Guilliams, M. *et al.* Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *J. Exp. Med.* **210**, 1977–1992 (2013).
- Epelman, S. *et al.* Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. *Immunity* **40**, 91–104 (2014).
- Hashimoto, D. *et al.* Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* **38**, 792–804 (2013).
- Hoeffel, G. *et al.* Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac-derived macrophages. *J. Exp. Med.* **209**, 1167–1181 (2012).
- Kumaravelu, P. *et al.* Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. *Development* **129**, 4891–4899 (2002).
- Bain, C. C. *et al.* Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nat. Immunol.* **15**, 929–937 (2014).
- Rantakari, P. *et al.* The endothelial protein PLVAP in lymphatics controls the entry of lymphocytes and antigens into lymph nodes. *Nat. Immunol.* **16**, 386–396 (2015).
- Haldar, M. *et al.* Heme-mediated SPI-C induction promotes monocyte differentiation into iron-recycling macrophages. *Cell* **156**, 1223–1234 (2014).
- Carlin, L. M. *et al.* Nr4a1-dependent Ly6C^{low} monocytes monitor endothelial cells and orchestrate their disposal. *Cell* **153**, 362–375 (2013).
- Stan, R. V. Endothelial stomatal and fenestral diaphragms in normal vessels and angiogenesis. *J. Cell. Mol. Med.* **11**, 621–643 (2007).
- Stan, R. V. *et al.* The diaphragms of fenestrated endothelia: gatekeepers of vascular permeability and blood composition. *Dev. Cell* **23**, 1203–1218 (2012).
- Tchaikovski, V., Fellbrich, G. & Waltenberger, J. The molecular basis of VEGFR-1 signal transduction pathways in primary human monocytes. *Arterioscler. Thromb. Vasc. Biol.* **28**, 322–328 (2008).
- Dejda, A. *et al.* Neuropilin-1 mediates myeloid cell chemoattraction and influences retinal neuroimmune crosstalk. *J. Clin. Invest.* **124**, 4807–4822 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the following people for their expert technical assistance: E.-L. Väänänen, R. Sjöroos, S. Mäki, M. Pohjansalo, S. Tyystjärvi and P. Laasola. We thank J. Lilja and G. Jacquemet for advice. We acknowledge the Cell Imaging Core at the Turku Centre for Biotechnology, The Finnish Microarray and Sequencing Centre and the Laboratory of Electron Microscopy in University of Turku. The research was supported by grants from the Academy of Finland (to E.P., J.I. and M.S.), the Juselius Foundation (to P.R. and M.S.), the Cancer Foundation (to M.S.), the South-Western Regional Fund of the Finnish Cultural Foundation and the Foundation of Turku University (to P.R.) and the Satakunta Regional Fund of the Finnish Cultural Foundation (to N.J.).

Author Contributions P.R. and N.J. contributed to the study design, and conducted most *in vivo* experiments and all FACS studies. E.L. and E.M. performed the whole-mount studies and the paraffin stainings, respectively. H.G. assisted with the *in vivo* experiments, and E.P. performed the ductal branching assays. J.I. planned, performed and analysed most biochemical experiments, and K.E. supervised and analysed the PLVAP-Fc generation and qPCR assays. K.A. conducted most of the confocal and all of the electron microscopy studies. M.S. conceived and supervised the study, planned experiments, analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (marko.salmi@utu.fi).

Reviewer Information Nature thanks D. Cheresh and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. *Plvap*^{tm1Salm} (*Plvap*^{−/−}), *Nt5e*^{tm1Lfr} (*Nt5e*^{−/−}) and *Aoc3*^{tm1Salm} (*Aoc3*^{−/−}) mice have been previously described^{24,31,32}. *Nt5e*^{−/−} and *Aoc3*^{−/−} mice manifest with altered leukocyte trafficking³³. We produced inducible *Plvap*^{F/F}; *CAGGCre-ER*TM mice by crossing conditional *Plvap*^{F/F} mice with a *CAGGCre-ER*TM (B6.Cg-Tg (CAG-cre/Esr1)^{5Amc/J} (stock 004682 from the Jackson Laboratory) deleter Cre mouse line, in which *Cre-ER* is ubiquitously expressed after tamoxifen injection³⁴. To achieve selective deletion of *Plvap* during embryonic development in LYVE-1⁺ endothelial cells, including yolk-sac endothelium and liver sinusoidal endothelial cells, but not the majority of other blood vessels³⁵, conditional *Plvap*^{F/F} mice were crossed with *Lyve1-Cre* mice (*Lyve1*^{tm1.1(EGFP/cre)Cys}, stock 012601 from the Jackson Laboratory³⁶) to generate *Plvap*^{F/F}; *Lyve1-Cre* mice. *Cav1*^{tm1Mls/J} (*Cav1*^{−/−}, stock 004585) mice, which lack all caveolae³⁷, were obtained from the Jackson Laboratory. C57BL/6J, C57BL/6N and BALB/c mice were purchased from Charles River and Janvier labs. F1 hybrid C57BL/6;129 (stock 101045) mouse strain was obtained from the Jackson Laboratory.

Both genders were used in the experiments (except in mammary gland analyses). Sex-matched wild-type littermate mice were used as controls in each experiment. Embryonic development was estimated considering the day of vaginal plug as embryonic age of 0.5 days (E0.5). The adult mice were 4–5 weeks old, since few *Plvap*^{−/−} mice survive till early adulthood^{24,28,38}. All animal experiments were approved by the Ethical Committee for Animal Experimentation in Finland. They were carried out in adherence with the rules and regulations of the Finnish Act on Animal Experimentation (497/2013) and in accordance to the 3R-principle under Animal License number 5587/04.10.07/2014.

Genotyping. Genotyping of *Plvap*^{−/−}, *Cav1*^{−/−}, *Nt5e*^{−/−} and *Aoc3*^{−/−} mice was performed according to protocols described previously^{24,31,32,37}. *Plvap*^{F/F}; *CAGGCre-ER*TM and *Plvap*^{F/F}; *Lyve1-Cre* mice genotyping was conducted using the following primers: primers A (3′-GTACATGCAACACCACTGAGC-5′) and B (3′-CCTTGACAGGTGATGTCTGC-5′) detect the wild-type *Plvap* allele (a 210-bp fragment) and the targeted *Plvap* allele (a 310-bp fragment; data not shown). Genotyping of *CAGGCre-ER*TM and *Lyve1-Cre* was done according to protocols described previously^{34,36}.

Isolation of embryonic and adult cells. Pregnant females were killed by carbon dioxide inhalation and cervical dislocation. Embryos from E10.5–E16.5 were dissected out from uterus and immersed in cold PBS (Invitrogen). The blood was collected after decapitation to heparin-containing tubes. Liver, lungs, spleen and brains were carefully dissected from the embryo and the yolk sac was collected. To obtain single-cell suspensions, the organs were incubated in Hank's buffered saline (HBS) containing 1 mg ml^{−1} collagenase D (Roche), 50 µg ml^{−1} DNase I (Sigma) at 37 °C in 5% CO₂ (30 min for liver, lung, spleen and brain and 2 h for yolk sac), and then passed through a 70-µm cell strainer. Erythrocytes were lysed from the blood and spleen samples as described²⁴. The brain cells were re-suspended in isotonic Percoll and the microglia were isolated as described⁶.

The cells from the adult tissues were isolated by the same method with some modifications. The blood was collected by a cardiac puncture into heparinized tubes. Lymph nodes, lung tissue and mammary fat pads were mechanically dissociated before a 60-min collagenase D and DNase I digestion. Livers were dissociated using Gentle MACS C-tube (Miltenyi Biotec) and immune cells were purified via OptiPrep density gradient centrifugation (Sigma D1556). The bone marrow was isolated by gently crushing the femurs before filtration. Lamina propria cells from the colon were isolated by an enzymatic digestion as described²³. Peritoneal cells were collected by flushing the peritoneal cavity with RPMI 1640 supplemented with 2% FBS and 5 IU ml^{−1} heparin.

Total leukocyte numbers in different organs were enumerated by determining the absolute numbers of viable cells in the cell suspensions by an automated cell counter (Cellometer Auto 2000, Nexcelcom) and the percentage of CD45⁺ cells (and of the various leukocyte subpopulations) by flow cytometry. Absolute leukocyte numbers in the blood were counted using an automated haemocytometer (VetScan HM5, Abaxis).

Yolk sac macrophage depletion. Pregnant heterozygous *Plvap*^{+/−} females were transiently treated with anti-CSF-1R monoclonal antibody (clone AFS98, Bio X Cell) or with the rat IgG2a isotype control (clone 2A3, Bio X Cell) at E6.5 using a single intraperitoneal injection (3 mg of antibodies in sterile PBS). This treatment prevents the development of yolk sac macrophages, but does not affect EMP development^{10,39}.

Flow cytometry and FACS. The fluorochrome-conjugated monoclonal antibodies (the antibody clones, fluorochromes, suppliers and catalogue numbers) against mouse molecules that were used for flow cytometry stains are listed in Supplementary Table 2. Before staining, the cell suspensions were incubated with purified anti-CD16/32 (clone 2.4G2, 553142 from Becton Dickinson) for 10 min on ice to block non-specific binding to Fc-receptors. Isotype-matched negative control antibodies conjugated to the appropriate fluorochromes were used

(Supplementary Table 2). All FACS analyses were run using an LSRFortessa flow cytometer (BD Biosciences) and analysed using FlowJo (TreeStar) software. The FACS gates used to define each leukocyte subpopulation in different organs and tissues of embryos and adult mice are shown in Extended Data Figs 1, 2b, d, g, 4b, 5c–e, 10d and Supplementary Table 1.

EMPs (CD41⁺CD45⁺c-Kit^{high}F4/80[−] cells) from E10.5 yolk sac, and yolk-sac-derived macrophages (CD11b⁺F4/80^{high} cells) and fetal liver-derived macrophages (CD11b⁺F4/80^{intermediate} cells) from E14.5 (wild-type and *Plvap*^{−/−}) and E16.5 (wild-type) fetal livers (mechanical dissociation without enzymatic digestions) were sorted from embryos using Sony SH800Z (100-µm nozzle, Sony Biotechnology) and FACS aria II (70-µm nozzle, Becton Dickinson) cell sorters. The purity of the isolated populations was >95%.

Cytology. Approximately 5,000 macrophages sorted from fetal livers (see above) were spun down onto microscopic slides using a cytospin centrifuge (Shandon cytospin III, Tecan). The cells were stained with Diff-Quick (REASTAIN), and photographed using Zeiss AxioVert 200M (Zeiss) using a Plan-Noefluar 40×/0.60 objective.

Immunohistochemistry and image analysis. Fetal livers from E12.5–E16.5 embryos were excised from the mice after decapitation. They were embedded in optimal cutting temperature (OCT) compound and snap-frozen. Cryostat sections (6 µm in thickness) were cut and fixed in ice-cold acetone. The sections were overlaid with the following antibodies: Alexa Fluor 488-conjugated rat anti-mouse F4/80 (MF48020, Invitrogen), allophycocyanin-conjugated rat anti-mouse CD31 (102510, BioLegend), rat monoclonal anti-mouse MECA32 (550563, Becton Dickinson), rat anti-mouse MAdCAM-1 (MECA-367; rat IgG2a, a gift from E. Butcher), rabbit polyclonal anti-mouse LYVE-1 (102-PA50AG or 103-PA50; Reliatech), rabbit polyclonal anti-mouse caveolin (SC-894; Santa Cruz) and rabbit polyclonal anti-VEGF (46154, Abcam). Alexa Fluor 647-conjugated goat anti-rat immunoglobulin (A21247, Life Technologies), Alexa Fluor 546-conjugated goat anti-rabbit immunoglobulin (A11035 and A11035, highly cross-absorbed, Invitrogen), Alexa Fluor 633-conjugated goat anti-rabbit immunoglobulin (highly cross-adsorbed A21071, Life Technologies) and Alexa Fluor 488-conjugated donkey anti-rat immunoglobulin (highly cross-adsorbed, A21208, Life Technologies) were used as secondary antibodies as appropriate. The sections were mounted in ProLong Gold with or without DAPI (4′,6-diamidino-2-phenylindole).

For visualization of the luminal location of PLVAP in fetal liver sinusoids at E12.5, wild-type C57BL/6N dams were killed at E12.5 and the embryos were excised from the uterine cavity but kept inside the yolk sac in warm PBS. Then 20 µg of unconjugated rat monoclonal anti-mouse MECA-32 (MECA32, custom-made, InVivo BioTech) or isotype-matched control antibody (rat IgG2a 553926 BD) was injected to umbilical and vitelline veins of the yolk sac. After 1 min, the embryos were decapitated, and the livers were collected and processed for *ex vivo* immunostaining. Acetone-fixed frozen sections were sequentially stained with Alexa Fluor 647-conjugated goat anti-rat immunoglobulin (A21247, Life Technologies) to detect the *in vivo* bound MECA-32, rabbit anti-mouse LYVE-1 (103-PA50, Reliatech) and Alexa Fluor 546-conjugated goat anti-rabbit immunoglobulin (A11035, Invitrogen). Preliminary analyses verified that species-specific second-stages antibodies showed no cross-reactivity with primary antibodies generated in the other species.

Whole-mount immunohistochemistry from optically cleared E8.5 and E10.5 yolk sac and AGM, E9.5 embryos and E14.5 fetal livers was done as described previously²⁴. Primary antibodies were rat monoclonal antibodies against mouse CD117 (c-Kit, 553352, Becton Dickinson), MECA-32 (PLVAP, 550563, Becton Dickinson), CD31 (550274, Becton Dickinson) and rabbit anti-mouse LYVE-1 (103-PA50, Reliatech). Alexa Fluor 488-conjugated donkey anti-rat immunoglobulin (A21208), Alexa Fluor 546-conjugated goat anti-rat immunoglobulin (A11081), Alexa Fluor 647-conjugated goat anti-rat immunoglobulin (A21247), and Alexa Fluor 633-conjugated goat anti-rabbit immunoglobulin (A21071) were used as secondary antibodies (all from Life Technologies). In AGM, c-Kit is expressed in HSCs, and CD31 in endothelial cells and HSCs⁴⁰.

A human fetal liver sample (pregnancy week 18) was cut, acetone-fixed and stained with monoclonal PAL-E (against human PLVAP; Ab8086, Abcam) and Alexa Fluor 488-conjugated goat anti-mouse immunoglobulin (highly cross-absorbed, A11029, Invitrogen). After staining, the sections were mounted in Prolong Gold with DAPI.

Images were acquired with a LSM 780 confocal microscope (Zeiss) using a c-Apochromat 40×/1.20 W Korr M27 objective or plan-apochromat 20×/0.8 objective (Fig. 4a and Extended Data Fig. 4g) and Zen 2010 software (Zeiss). Using pinhole adjustments, a slice thickness of 4.6 µm and 1.2 µm was used for 20× and 40× objectives, respectively. A background subtraction was used for all images. In certain images, the brightness was linearly changed and noise was reduced using mean filter in ImageJ software. Brightness adjustments and noise reductions were always applied equally to images captured from wild-type and PLVAP-deficient

mice. Splenic F4/80^{high} cells were quantified by thresholding the images so that only the F4/80^{high} cells remained visible (the thresholding was applied equally to images captured from wild-type and PLVAP-deficient mice). Thereafter, white pulp areas were excluded (based on MacCAM-1 staining in marginal zone⁴¹) and the area fraction in the red pulp containing the F4/80^{high} cells was measured using ImageJ software. A spleen area of at least 2.1 mm² per mouse was analysed.

Z-stacks and images (Extended Data Fig. 4a E8.5 yolk sac) were acquired from the optically cleared samples using a 3i Spinning Disk confocal microscope (Intelligent Imaging Innovations) with a plan-apochromat 20×/0.8, 10×/0.45 or LD c-apochromat 40×/1.1 W objective. Background subtractions, linear brightness adjustments and mean filter noise reductions were done using ImageJ software. To produce maximum projections with SlideBook 6 software (Intelligent Imaging Innovations, Inc.), 15–87 sections with slice thickness of 0.63 µm were used (Extended Data Fig. 4a E10.5 yolk sac and 4g). 3D reconstruction was generated from Z-stacks (44 slices with thickness of 2.34 µm (MECA-32 in Supplementary Video 1), 78 slices with slice thickness of 2.34 µm (CD31 in Supplementary Video 1), and 117 slices with thickness of 0.43 µm (MECA-32 in Supplementary Video 2)) and converted to AVI file format with Imaris 8.0 software (Bitplane).

Histology and iron staining. Formalin-fixed, paraffin-embedded sections of livers from E11.5–E16.5 wild-type and *Plvap*^{−/−} mice were cut, deparaffinized and subjected to heat-mediated antigen retrieval in EDTA buffer (Dako S2367). Endogenous peroxidase was quenched with 3% H₂O₂, non-specific immunoglobulin binding was blocked with rabbit serum, and endogenous biotin and avidin were blocked using DakoCytomation Biotin blocking system (Dako, X0590). The liver sections were incubated overnight at 4 °C with the primary antibody (MECA-32, 1 µg ml^{−1} in PBS). A secondary antibody (biotinylated anti-rat immunoglobulin) was incubated for 30 min, and then biotin-avidin complexes were formed with using Vectastain ABC kit (PK-6100, Vector Laboratories). Liquid DAB+ substrate Chromogen System (Dako K3468) was used to oxidize and detect the peroxidase complexes. Finally, samples were stained with haematoxylin, dehydrated and mounted.

Macrophage-dependent iron recycling was studied in spleen and liver by measuring accumulation of ferric ion^{42,43}. For Fe³⁺ stains, the spleens and livers from 5-week-old wild-type and *Plvap*^{−/−} mice were fixed with 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.0), embedded in paraffin and sectioned. The detection of ferric iron was accomplished using the Prussian blue histological staining method, as described previously^{25,43}. The slides were counter-stained in nuclear Fast Red for 5 min. The slides were analysed using a Panoramic 250 Flash II slide-scanner (3D Histech). In spleen tissue, the white pulp areas were excluded and the red pulp areas containing the blue-stained Fe³⁺-containing cells were analysed using image thresholding. Area fractions were measured using ImageJ software. A spleen area of at least 1.5 mm² per mouse was analysed. In livers, whole sections (at least 21.1 mm² per mouse) were analysed.

Analysis of ductal morphogenesis in mammary glands. The fourth mammary gland was dissected from 4.5-week-old wild-type and *Plvap*^{−/−} mice, and left to adhere to the object glass. The tissue was fixed by submerging in Carnoy's medium (60% ethanol, 30% chloroform, 10% acetic acid) overnight at 4 °C, and rehydrated in decreasing ethanol concentration series. The slides were then stained with carmine alum (0.2% carmine, 0.5% aluminium potassium sulfate dodecahydrate) overnight at room temperature, dehydrated, cleared in xylene for 2–3 days, and mounted in DPX Mountant (Sigma). The samples were imaged with Zeiss SterEO Lumar V12 stereo microscope using NeoLumar 0.8× objective and Zeiss AxioCam ICc3 colour camera. Several images were automatically combined into a mosaic picture using Adobe Photoshop. The area covered by the ductal tree, and the number of ductal branches was tracked manually and quantified using ImageJ with 'Skeletonize2D/3D' and 'AnalyzeSkeleton' plugins.

Electron microscopy. Livers from E12.5, E14.5 and E16.5 wild-type and *Plvap*^{−/−} embryos were collected and fixed in 5% glutaraldehyde in 0.16 M s-collidine buffer, pH 7.4. The samples were post-fixed for 2 h with 2% OsO₄ containing 3% potassium ferrocyanide, dehydrated with a series of increasing ethanol concentrations (70%, 96% and twice at 100%) and embedded in 45359 Fluka Epoxy Embedding Medium kit. 70-nm sections were cut with an ultramicrotome, and stained with 1% uranyl acetate and 0.3% lead citrate. The sections were examined with a JEOL JEM-1400 Plus transmission electron microscope.

Colony-forming assays. After isolation, 5,000 cells from the yolk sac of E10.5, liver of E12.5 embryos and adult bone marrow of wild-type and *Plvap*^{−/−} mice were seeded in 1 ml of M3434 Methocult medium (Stem Cell Technologies) into 35-mm culture dishes in duplicates. After a 7-day culture at 37 °C with 5% CO₂ the number of colonies was counted, as described⁴⁴.

Quantitative PCR. Total RNA was isolated from fetal livers of wild-type and PLVAP-deficient mice using the Nucleo-Spin RNA kit (Macherey–Nagel) and from the sorted EMP and macrophages (see above) using the RNeasy Plus Micro kit (Qiagen). The RNA was reverse-transcribed to cDNA with SuperScript VILO cDNA

Synthesis kit (ThermoFisher Scientific) according to the manufacturers' instructions. Quantitative PCR (qPCR) was carried out using Taqman Gene Expression Assays (ThermoFisher Scientific) for *Plvap* (Mm00453379_m1; target gene), *Lyve1* (Mm00475056_m1; target gene) and *Actb* (Mm00607939_s1; control gene). The expression of reported signature transcripts⁷ enriched in yolk-sac-derived F4/80^{high} macrophages (*Cx3cr1* (Mm00438354_m1), *Mrc1* (Mm00485148_m1), *Adgre1* (Mm00802529_m1, also known as *Emr1* or *F4/80*), and in F4/80^{intermediate} fetal liver-derived monocytes (*Itgam* (Mm00434455_m1), *Gata2* (Mm00492301_m1), *Flt3* (Mm00439016_m1) and *Ccr2* (Mm04207877_m1)) in E16.5 fetal liver of wild-type mice was also analysed by quantitative qPCR. The reactions were run using the 7900HT Fast Real-Time PCR System (Applied Biosystems/ ThermoFisher Scientific) or QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems/ ThermoFisher Scientific) at the Finnish Microarray and Sequencing Centre (FMSC), Turku Centre for Biotechnology, Turku, Finland. Relative expression levels were calculated using Sequence Detection System (SDS) Software v2.4.1, QuantStudio 12K Flex software, and DataAssist software (all from Applied Biosystems/ThermoFisher Scientific). The results were presented as a percentage of the control gene mRNA level from the same samples.

Interaction studies. A PLVAP–Fc fusion protein expressing the extracellular domain of mouse PLVAP fused to human IgG2 Fc-tail was generated (Extended Data Fig. 9c). The extracellular domain (amino-acid residues 48–438) was PCR-cloned from a full-length cDNA clone for mouse PLVAP (MR206983, Origene) using primers introducing EcoRV and NheI digestion sites. The PCR reaction was carried out using Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific). The amplified fragment was purified and annealed to EcoRV and NheI digested pFUSEN-hG2Fc vector (InvivoGen) designed for the production of Fc-chimaeras from type 2 membrane proteins. The intactness of the construct was verified by sequencing, and its reactivity with anti-PLVAP antibody MECA-32 using immunoblotting. The expression plasmid was transfected into HEK293-EBNA cells (CRL-10852, from ATCC) using lipofection (Lipofectamine, Invitrogen), the cells were cultured for 2–3 days in serum-free medium (Pro293A-CDM, Bio-Whittaker). A CD4–Fc chimaera⁴⁵ was used as a control.

For heparin-affinity pull-down assays, agarose beads coupled to heparin (Sigma) or streptavidin (negative-control beads, from GE Healthcare) were washed, and blocked with TBS (pH 7.2) containing 1% BSA. The beads were rocked with clarified 0.5% NP-40 total protein lysates from E14.5 wild-type livers for 2 h at 4 °C in TBS containing 1% BSA. Alternatively, PLVAP–Fc and CD4–Fc fusion proteins were applied to the heparin and control beads. After washing with TBS containing 0.3% NP-40, the bead-bound molecules were eluted in Laemmli's sample buffer, and subjected to SDS–PAGE separation. In certain experiments, the heparin and control beads were incubated with the fusion proteins and, after washing, the same volumes of the beads were eluted directly in Laemmli's sample buffer or in 1.0 M NaCl, and the eluted proteins from the supernatants were submitted to SDS–PAGE. In other specificity control experiments, the binding of PLVAP–Fc fusion protein to heparin-beads was analysed in the absence and presence of 100 µg of fibronectin (F1141, Sigma), which binds to heparin⁴⁶, or 100 µg of collagen (C8919, Sigma). After transfer to nitrocellulose membranes, the bound molecules were visualized using immunoblotting with a horseradish-peroxidase-conjugated anti-human IgG antibody (81-7120, Invitrogen; for the chimaeras) or with anti-PLVAP (MECA-32, BioXCell), anti-neuropilin-1 (AF56615, R&D Systems), and anti-VEGF (sc-152, Santa Cruz) antibodies followed by appropriate HRP-conjugated second-stage reagents (for the liver lysates) using ECL detection.

For far-western assays, recombinant mouse neuropilin-1 (R&D Systems, 5994-N1) and recombinant mouse VEGF164 (R&D Systems, 493-MV) were spotted onto filters. Both neuropilin-1 and VEGF are known heparin-binding proteins^{47,48}. The PLVAP–Fc chimaera (in TBS containing 5% BSA) was allowed to bind to the immobilized proteins in the presence or absence of 50 IU heparin (stock 5,000 IU ml^{−1}; Leo Pharma) for 2 h. The bound chimaera was visualized using HRP-conjugated anti-human IgG antibody and ECL.

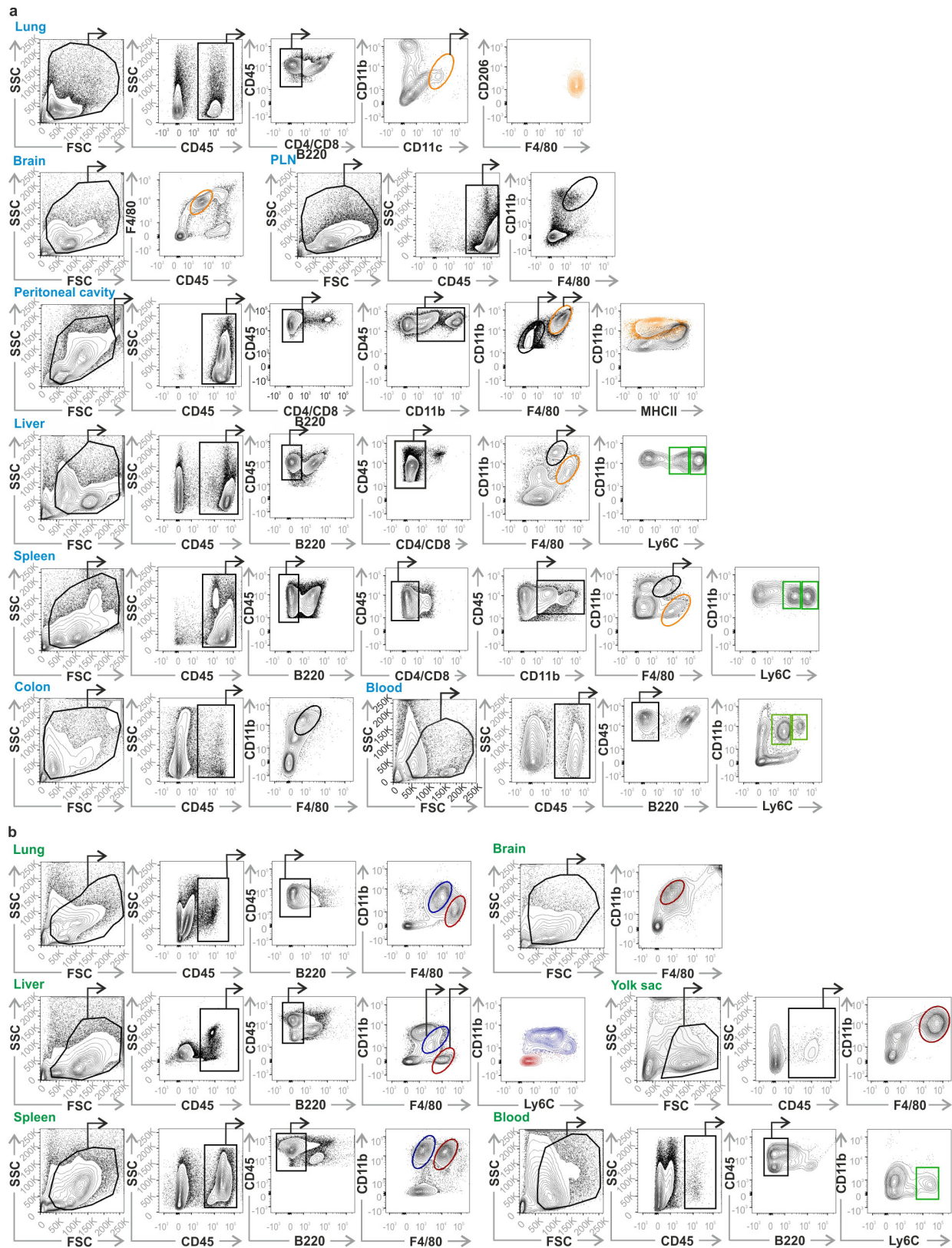
Detection of VEGF interaction with PLVAP *in situ* in E14.5 fetal livers was performed using a proximity ligation assay (PLA)⁴⁹. In brief, primary antibodies were rabbit anti-VEGF (46154, Abcam) or rabbit anti-GFP (A11122, Molecular Probes; as a negative control), and they were detected by Duolink *in situ* PLA probe anti-rabbit PLUS (DUO92002, Sigma), and rat anti-PLVAP antibody (MECA-32) was directly conjugated to MINUS PLA probe using Duolink *in situ* problemaker MINUS kit (DUO92010, Sigma). After ligation and amplification, the probes were detected using Detection reagent red (DUO92008, Sigma). During the amplification step, Alexa Fluor 488-conjugated donkey anti-rat IgG (A11035, Invitrogen) was added to detect MECA-32. The samples were stained with DAPI and mounted in Mowiol. Images for PLA were acquired using a 3i Spinning Disk confocal microscope (Intelligent Imaging Innovations) with a plan-apochromat 63×/1.4 oil objective and SlideBook 6 software (Intelligent Imaging Innovations). Background subtractions and linear brightness adjustments were performed using

ImageJ. Adjustments were applied equally to images captured from control and anti-VEGF antibody PLVAP PLA stains.

For co-immunoprecipitation assays, freshly isolated E14.5 livers of wild-type mice were briefly lysed in a buffer containing 1% NP-40, 150 mM NaCl, 20 mM HEPES (pH 7.5), 2 mM MgCl₂, 2 mM CaCl₂, PhosSTOP and Protease inhibitor cocktail (both Roche). After clarification by centrifugation, the supernatants were incubated with a rabbit anti-VEGFA antibody (or with a negative control rabbit antibody) for 5 h at 4 °C. Protein G beads (blocked with 1% BSA) were then added for 1 h at 4 °C, and thereafter the beads were washed 3 times with the lysis buffer. The bound proteins were eluted in non-reducing Laemmli's sample buffer, separated in SDS-PAGE and immunoblotted for VEGF and PLVAP using IRDye-conjugated second-stage reagents and Odyssey imager.

Statistics. Sample size was empirically determined based on pilot analyses and previous literature. Adult wild-type and *Plvap*^{-/-} littermates were allocated to experimental groups without specific randomization methods because comparisons involved mice of distinct genotypes. The investigators were blinded to the genotype of the embryos during the experimental procedures. Numerical data are given as mean ± s.e.m. Comparisons between genotypes were performed using Mann-Whitney *U*-test. SAS 9.4 statistical software and GraphPad Prism software v6 were used for statistical analysis. *P* < 0.05 was considered to be statistically significant. Each data point (values provided in Source Data) is obtained from a different embryo or mouse, and thus all numeric data and statistical analyses are derived from biological replicates.

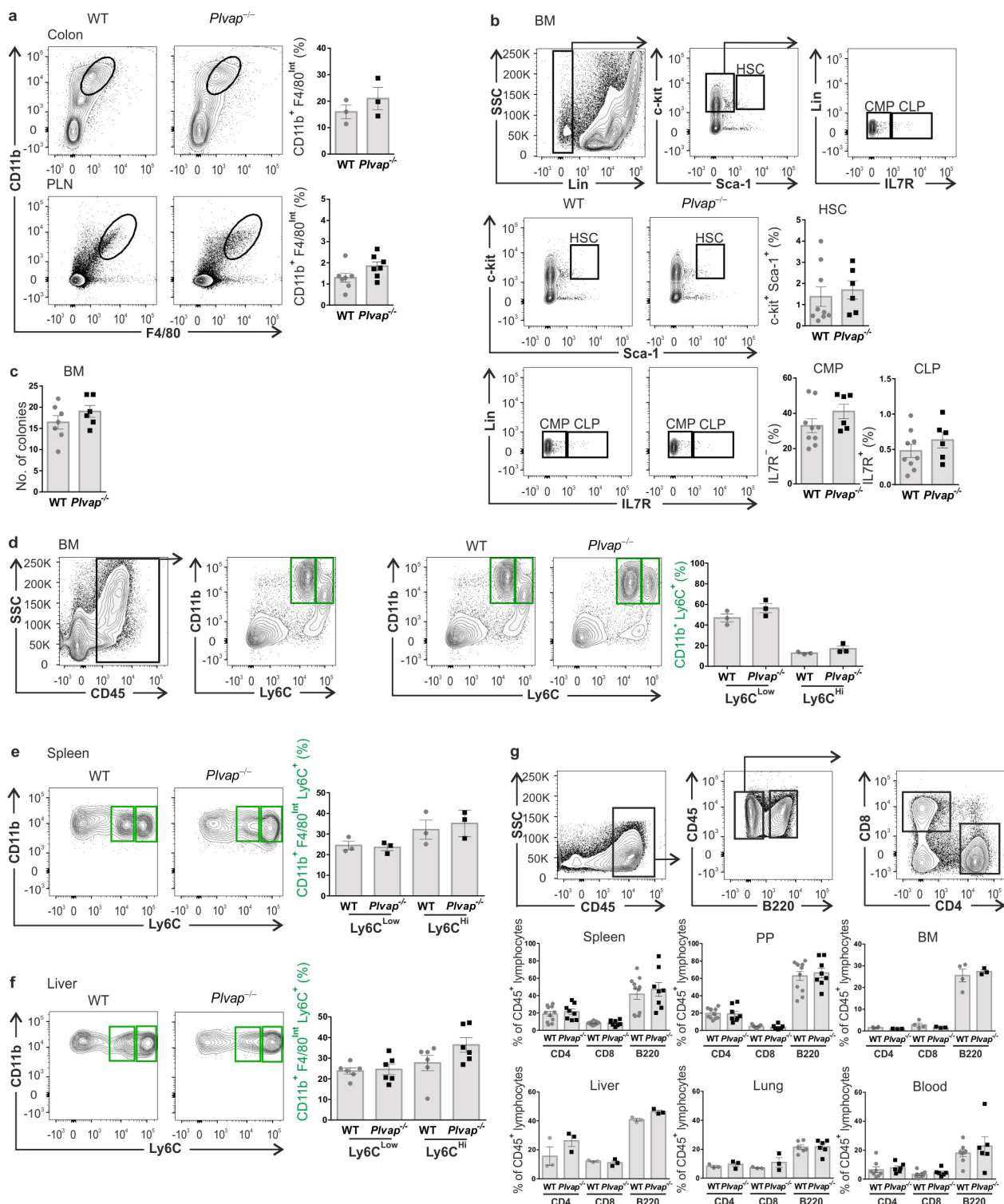
31. Thompson, L. F. *et al.* Crucial role for ecto-5'-nucleotidase (CD73) in vascular leakage during hypoxia. *J. Exp. Med.* **200**, 1395–1405 (2004).
32. Stolen, C. M. *et al.* Absence of the endothelial oxidase AOC3 leads to abnormal leukocyte traffic in vivo. *Immunity* **22**, 105–115 (2005).
33. Salmi, M. & Jalkanen, S. Cell-surface enzymes in control of leukocyte trafficking. *Nat. Rev. Immunol.* **5**, 760–771 (2005).
34. Hayashi, S. & McMahon, A. P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev. Biol.* **244**, 305–318 (2002).
35. Gordon, E. J., Gale, N. W. & Harvey, N. L. Expression of the hyaluronan receptor LYVE-1 is not restricted to the lymphatic vasculature; LYVE-1 is also expressed on embryonic blood vessels. *Dev. Dyn.* **237**, 1901–1909 (2008).
36. Pham, T. H. *et al.* Lymphatic endothelial cell sphingosine kinase activity is required for lymphocyte egress and lymphatic patterning. *J. Exp. Med.* **207**, 17–27 (2010).
37. Razani, B. *et al.* Caveolin-1 null mice are viable but show evidence of hyperproliferative and vascular abnormalities. *J. Biol. Chem.* **276**, 38121–38138 (2001).
38. Herrnberger, L. *et al.* Lack of endothelial diaphragms in fenestrae and caveolae of mutant Plvap-deficient mice. *Histochem. Cell Biol.* **138**, 709–724 (2012).
39. Squarzone, P. *et al.* Microglia modulate wiring of the embryonic forebrain. *Cell Reports* **8**, 1271–1279 (2014).
40. Clarke, R. L. *et al.* The expression of Sox17 identifies and regulates haemogenic endothelium. *Nat. Cell Biol.* **15**, 502–510 (2013).
41. Kraal, G., Schornagel, K., Streeter, P. R., Holzmann, B. & Butcher, E. C. Expression of the mucosal vascular addressin, MAdCAM-1, on sinus-lining cells in the spleen. *Am. J. Pathol.* **147**, 763–771 (1995).
42. Ganz, T. Macrophages and systemic iron homeostasis. *J. Innate Immun.* **4**, 446–453 (2012).
43. Kohyama, M. *et al.* Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. *Nature* **457**, 318–321 (2009).
44. Swiers, G. *et al.* Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat. Commun.* **4**, 2924 (2013).
45. Salmi, M., Karikoski, M., Elima, K., Rantakari, P. & Jalkanen, S. CD44 binds to macrophage mannose receptor on lymphatic endothelium and supports lymphocyte migration via afferent lymphatics. *Circ. Res.* **112**, 1577–1582 (2013).
46. Hayashi, M., Schlesinger, D. H., Kennedy, D. W. & Yamada, K. M. Isolation and characterization of a heparin-binding domain of cellular fibronectin. *J. Biol. Chem.* **255**, 10017–10020 (1980).
47. Vander Kooi, C. W. *et al.* Structural basis for ligand and heparin binding to neuropilin B domains. *Proc. Natl Acad. Sci. USA* **104**, 6152–6157 (2007).
48. Ferrara, N. & Henzel, W. J. Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. *Biochem. Biophys. Res. Commun.* **161**, 851–858 (1989).
49. Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).



Extended Data Figure 1 | See next page for caption.

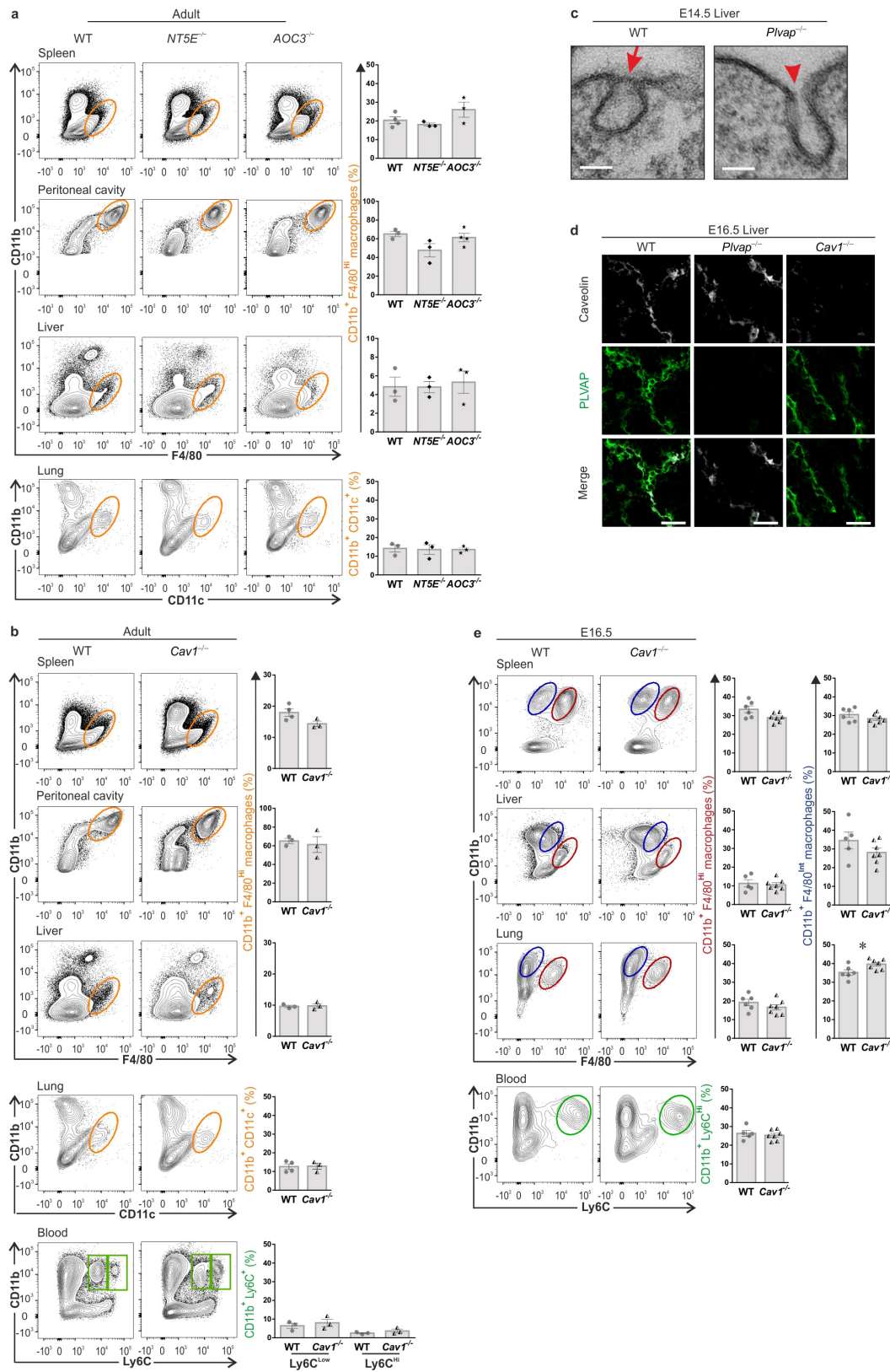
Extended Data Figure 1 | Full gating strategies for leukocyte subpopulations. a, b, Gating strategies of adult (a) and fetal (b) macrophage and monocyte populations in the indicated tissues. The colour code of the final gates is the same as in Fig. 1 (orange, embryonic-derived macrophages in adults; black, bone-marrow-derived macrophages in adults; red, yolk-sac-derived macrophages in embryos; blue, fetal liver monocyte-derived macrophages in embryos; green, monocytes). The rightmost panels in adult lung (CD206 versus F4/80), adult peritoneal cavity (CD11b versus MHCII) and fetal liver (CD11b versus Ly6C) are validation stains for the indicated populations (not used for gating). The gating strategies for the other studied leukocyte populations are shown

in Fig. 2c and Extended Data Fig. 5c (fetal liver macrophage–dendritic cell precursors, myeloid precursors, common monocyte progenitors, Ly6C⁺ and Ly6C[−] monocytes), Extended Data Fig. 2b (HSCs, common myeloid progenitors and common lymphoid progenitors in bone marrow), Extended Data Fig. 2d (Ly6C^{low} and Ly6C^{high} bone marrow monocytes), Extended Data Fig. 2g (CD4⁺, CD8⁺ and B220⁺ lymphocytes in the adult organs), Extended Data Fig. 4b (EMPs and macrophages in the yolk sac), Extended Data Fig. 5d, e (EMPs and HSCs in fetal liver) and Extended Data Fig. 10d (mammary gland leukocytes). The rest of the validation gates are listed in Supplementary Table 1.



Extended Data Figure 2 | Selective impairment in the accumulation of embryonic-derived tissue-resident macrophages in *Plvap*^{-/-} mice. **a, b,** Flow cytometry analyses of adult bone-marrow-derived *CD11b*⁺*F4/80*^{intermediate} tissue-resident macrophages (the black gate) in the colon and peripheral lymph nodes (PLN) (**a**), and HSCs (*Lin*⁻*c-Kit*⁺*Sca-1*⁺ cells), common myeloid (*Lin*⁻*c-Kit*⁺*Sca-1*^{low}*IL7R*⁺; CMP) and common lymphoid (*Lin*⁻*c-Kit*⁺*Sca-1*^{low}*IL7R*⁺; CLP) progenitor cells in the bone marrow (BM) (**b**). **c,** Colony-forming assays on macrophage colony stimulating factor (M-CSF)-supplemented soft agar from bone marrow. **d–g,** Flow cytometry analyses of inflammatory (*CD11b*⁺*Ly6C*^{high}) and patrolling (*CD11b*⁺*Ly6C*^{low}) monocytes in the bone marrow (**d**, green gates), recently entered tissue monocytes (*CD11b*⁺*F4/80*^{intermediate}*Ly6C*^{high}

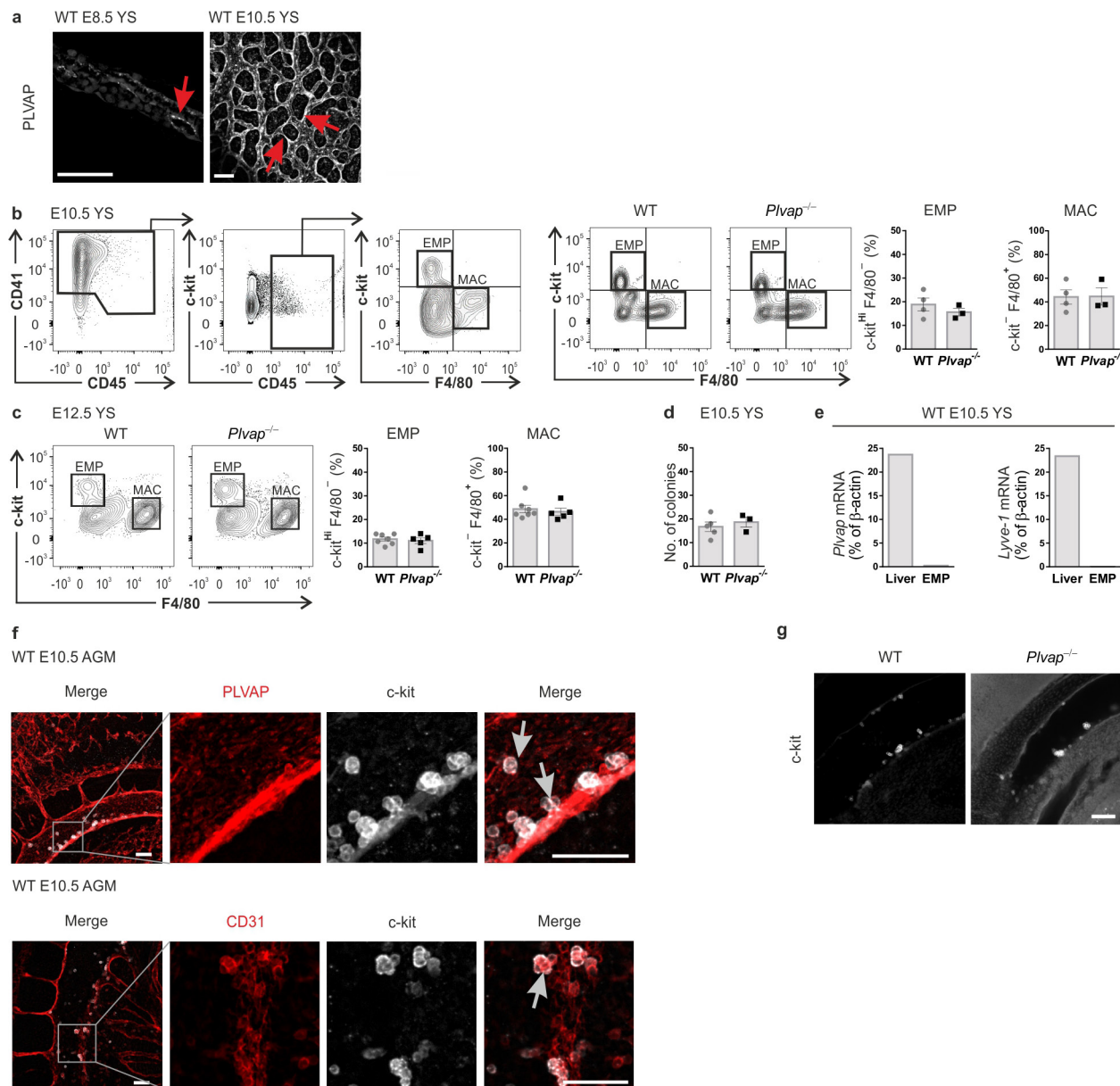
and *CD11b*⁺*F4/80*^{intermediate}*Ly6C*^{low} cells, green gates) in the spleen (**e**) and liver (**f**) and *CD4*⁺, *CD8*⁺ and *B220*⁺ lymphocytes in the spleen, Peyer's patches (PP), bone marrow (BM), liver, lung and blood (**g**) of adult mice. The flow cytometry data are shown as frequency of live-gated *CD45*⁺ cells (**a, d**), of live-gated *Lin*⁻ cells (**b**, for HSC), of live-gated *Lin*⁻*c-Kit*⁺*Sca-1*^{low} cells (**b**, for CMP and CLP), of live-gated *CD45*⁺*B220*⁻*CD4*⁻*CD8*⁻*CD11b*⁺*F4/80*^{intermediate} cells (**e, f**), of live-gated *CD45*⁺ cells (**g**, for B cells) and of live-gated *CD45*⁺*B220*⁻ cells (**g**, for *CD4* and *CD8* T cells). Each dot represents one mouse (pooled from 2–5 independent experiments, see Source Data), data are mean ± s.e.m. for each group.



Extended Data Figure 3 | See next page for caption.

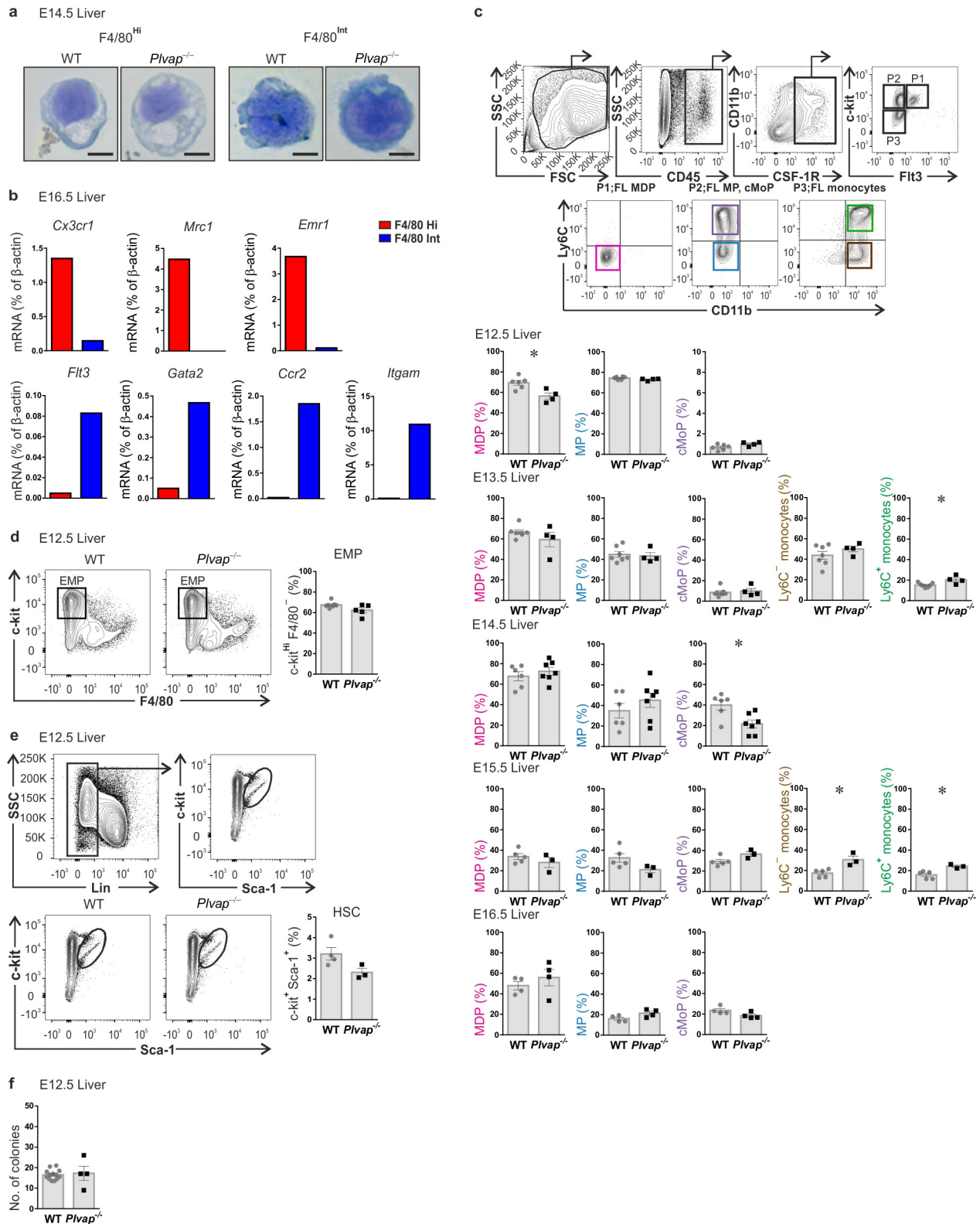
Extended Data Figure 3 | Normal seeding of embryonic-derived macrophages in *Nt5e*^{-/-}, *Aoc3*^{-/-} and *Cav1*^{-/-} mice. **a, b,** Flow cytometry analyses of embryonic-derived CD11b⁺F4/80^{high} tissue-resident macrophages (orange gates) in the spleen, peritoneal cavity and liver, and of embryonic-derived CD11b⁺CD11c^{high} macrophages (orange gates) in the lungs of adult wild-type, *Nt5e*^{-/-} and *Aoc3*^{-/-} mice (**a**), and in wild-type and *Cav1*^{-/-} mice (**b**). **c,** Electron micrographs of caveola in the fetal liver sinusoidal endothelium. Red arrow, a diaphragm-containing caveola; red arrowhead, a caveola without the diaphragm. **d,** Immunofluorescent stains of livers of wild-type, *Cav1*^{-/-} and *Plvap*^{-/-} mice at E16.5 for PLVAP and caveolin. **e,** Flow cytometry analyses

of yolk-sac-derived (CD11b⁺F4/80^{high}; red gates) and fetal liver-derived (CD11b⁺F4/80^{intermediate}; blue gates) macrophages, and of CD11b⁺Ly6C^{high} monocytes (green gate) in the blood in E16.5 wild-type and *Cav1*^{-/-} mice. Shown are representative images (*n* = 2 (**c**) and *n* = 3 (**d**) biological replicates from 4 (**d**) independent stains). Scale bars, 50 nm (**c**) and 20 μm (**d**). The flow cytometry data are shown as frequency of live-gated CD45⁺B220⁻CD4⁻CD8⁻ cells (**a, b**, adult tissues), and of CD45⁺B220⁻ (**b**, adult blood and **e**). Each dot represents one mouse or embryo (pooled from 2 independent experiments and from 2–3 litters (**e**), see Source Data), data are mean ± s.e.m. for each group (**P* < 0.05 by Mann–Whitney U-test).



Extended Data Figure 4 | EMP and macrophage accumulation in the yolk sac and c-Kit⁺ cell accumulation in the AGM are intact in *Plvap*^{-/-} mice. **a**, Immunohistochemical analyses of PLVAP expression in whole-mounts of wild-type yolk sac at E8.5 and E10.5. Red arrows, PLVAP⁺ vascular endothelium. **b**, **c**, Flow cytometry analyses of live-gated CD41⁺CD45⁺c-Kit^{high}F4/80⁻ EMP and CD45⁺c-Kit^{high}F4/80⁺ macrophages (MAC) at E10.5 (**b**), and at E12.5 (**c**) in the yolk sac. **d**, Colony-forming assays on M-CSF-supplemented soft agar from E10.5 yolk sac. **e**, qPCR analyses of *Plvap* and *Lyve1* expression in EMP of yolk sac in E10.5 wild-type mice (cells pooled from 30 embryos). Liver denotes

mRNA from whole E12.5 fetal liver (a positive control). **f**, Whole-mount immunofluorescent stains of c-Kit, CD31 and PLVAP in the AGM region of E10.5 wild-type mice. CD31 is expressed in endothelial cells and HSC. White arrows, representative c-Kit⁺ cells. **g**, Whole-mount immunostains of c-Kit in the AGM region of E10.5 wild-type and *Plvap*^{-/-} mice. Shown are representative images (**a**, **f**, **g**, $n = 3$ biological replicates from 2 independent stains). Scale bars, 50 μ m (**a**, **f**, **g**). Each dot represents one embryo (pooled from 2–3 independent experiments and from 2–3 litters (**b**, **c**, **d**), see Source Data), data are mean \pm s.e.m. for each group.

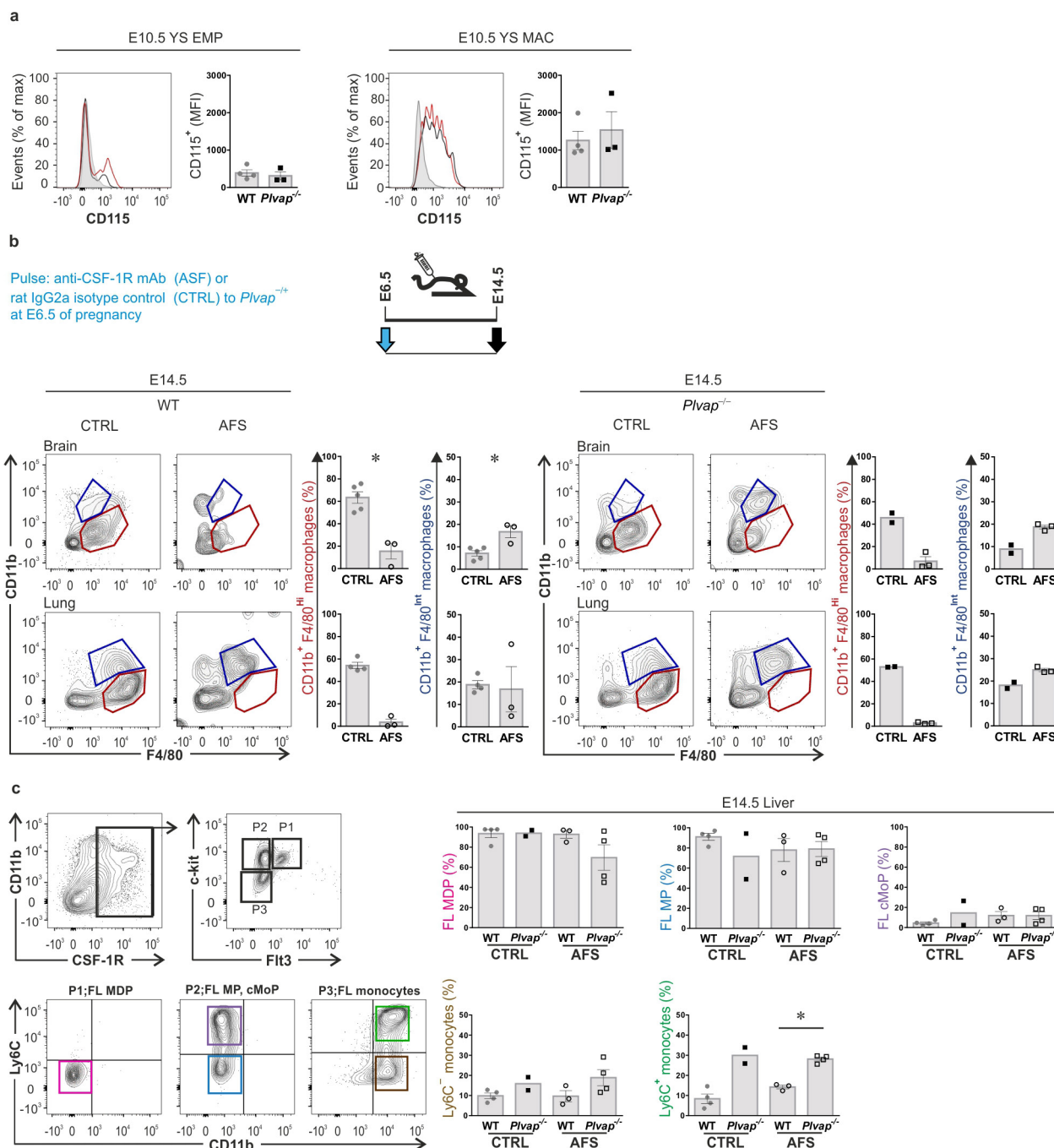


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Entry of monocyte progenitors to fetal liver and their differentiation to monocytes is PLVAP-independent.

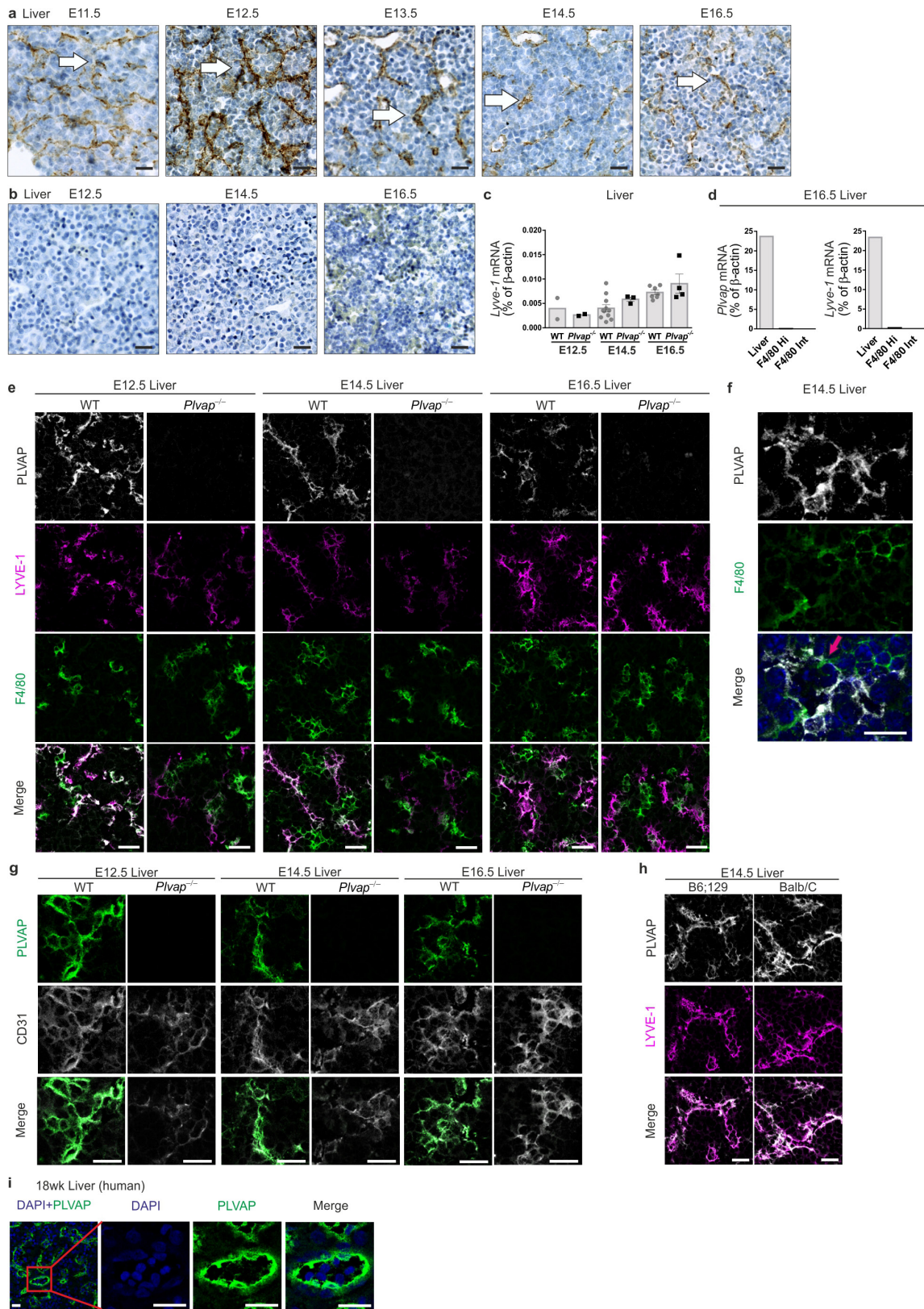
a, Cytospin stains of sorted $CD11b^{+}F4/80^{high}$ yolk-sac-derived macrophage-like cells and $CD11b^{+}F4/80^{intermediate}$ fetal liver-derived monocyte-like cells from E14.5 fetal livers of wild-type and *Plvap*^{-/-} mice (representative images from 4 embryos per genotype). Scale bars, 10 μ m. **b**, qPCR analyses of sorted $CD11b^{+}F4/80^{high}$ yolk-sac-derived macrophage-like cells (F4/80 Hi) and $CD11b^{+}F4/80^{intermediate}$ fetal liver-derived monocyte-like cells (F4/80 Int) isolated from E16.5 livers of wild-type mice for cell-type signature genes *Cx3cr1*, *Mrc1*, *Emr1*, *Flt3*, *Gata2*, *Ccr2*, and *Itgam* (cells pooled from 10 embryos). **c**, The gating strategy to identify fetal liver $CD11b^{-}CSF-1R^{+}c-Kit^{+}Flt-3^{+}Ly6C^{-}$ macrophage-dendritic cell precursors (MDP, first P1, then the pink gate), $CD11b^{-}CSF-1R^{+}c-Kit^{+}Flt-3^{-}Ly6C^{-}$ fetal liver myeloid precursors (MP, first P2, then the blue gate), and $CD11b^{-}CSF-1R^{+}c-Kit^{+}Flt-3^{-}Ly6C^{+}$ common

monocyte progenitors (cMoP, first P2, then the violet gate) and their enumeration at E12.5, E13.5, E14.5, E15.5 and E16.5. The gating strategy to identify fetal liver $CD11b^{+}CSF-1R^{+}c-Kit^{-}Flt-3^{-}Ly6C^{-}$ monocytes ($Ly6C^{-}$ monocytes, first P3, then the brown gate) and $CD11b^{+}CSF-1R^{+}c-Kit^{-}Flt-3^{-}Ly6C^{+}$ monocytes ($Ly6C^{+}$ monocytes, first P3 then the green gate), and their enumeration at E13.5 and E15.5 livers is also shown (the quantification of these two monocyte types at E12.5, E14.5 and E16.5 livers are shown in Fig. 2d). The flow cytometry data are shown as frequency of live-gated $CD45^{+}$ cells. **d**, Flow cytometry analyses of live-gated $CD41^{+}CD45^{+}c-Kit^{high}F4/80^{-}$ EMP cells at E12.5 in the livers. **e**, Flow cytometry analyses of live-gated $Lin^{-}c-Kit^{+}Sca-1^{+}$ HSC at E12.5 in the livers. **f**, Colony forming assays on M-CSF supplemented soft agar from E12.5 liver single-cell suspensions. Each dot represents one embryo (2–3 independent experiments and from 2–3 litters (**c–f**), see Source Data), data are mean \pm s.e.m. for each group (* $P < 0.05$ by Mann–Whitney U-test).



Extended Data Figure 6 | Deletion of yolk-sac-derived, but not EMP and fetal liver monocyte-derived macrophages by an anti-CSF-1R antibody injection to E6.5 pregnant mice. a, Flow cytometry analyses of CSF-1R (CD115) expression on live-gated CD41⁺CD45⁺c-Kit^{high}F4/80^{int} early EMPs in E10.5 yolk sac (the two panels on the left) and on live-gated CD45⁺c-Kit⁺F4/80⁺ macrophages (MAC) in E10.5 yolk sac (the two panels on the right) in wild-type and *Plvap*^{-/-} mice. Black histograms represent wild-type mice, red histograms represent *Plvap*^{-/-} mice, grey histograms show isotype-matched negative-control antibodies. MFI, mean fluorescence intensity. **b,** The anti-CSF-1R (AFS) and control antibody (CTRL) treatment strategy and flow cytometry analyses of yolk-sac-derived, CD11b⁺F4/80^{high} macrophages (red gates) and fetal liver-derived CD11b⁺F4/80^{intermediate} macrophages (blue gates) at E14.5 in the brain and lung of wild-type and *Plvap*^{-/-} embryos after a treatment of

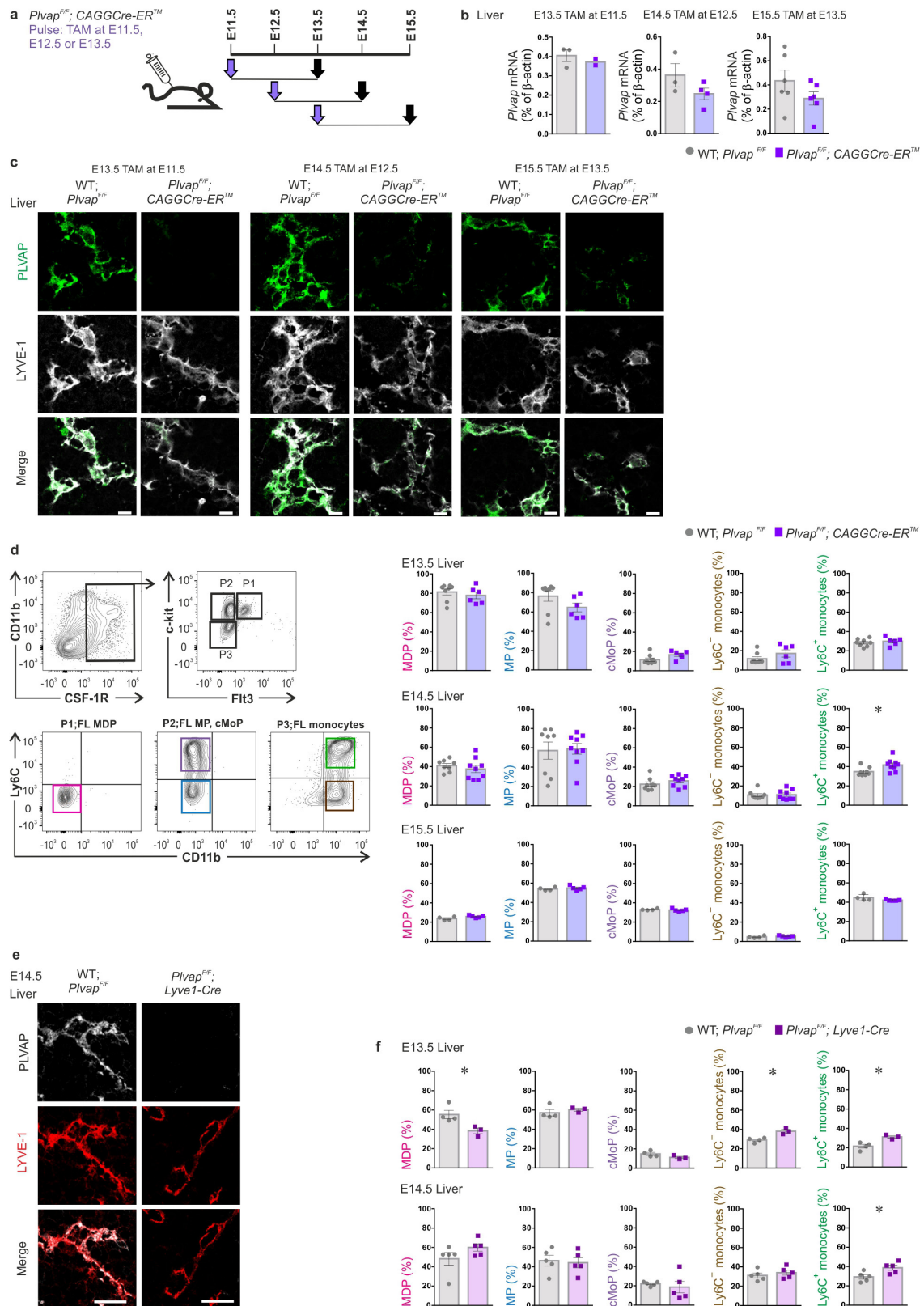
E6.5 pregnant *Plvap*^{+/-} mice with a single dose of anti-CSF-1R antibody (AFS) or isotype-matched control antibody. **c,** The gating strategy and flow cytometry analyses of the frequencies of fetal liver macrophage-dendritic cell precursors, myeloid precursors, common monocyte progenitors, Ly6C⁺ and Ly6C⁻ monocytes (gates as in Extended Data Fig. 5c and Supplementary Table 1) at E14.5 in the wild-type and *Plvap*^{-/-} embryos of *Plvap*^{+/-} dams, which were treated with the anti-CSF-1R antibody or isotype-matched control antibody at E6.5 of pregnancy. The flow cytometry data are shown as frequency of live-gated CD45⁺ cells (**a**), of live-gated CD45⁺B220⁻ cells (**b**, lung), and of live-gated CD45⁺ cells (**b**, brain and **c**). Each dot represents one embryo (pooled from 2–3 independent experiments from 2–4 litters (**a–c**), see Source Data), data are mean ± s.e.m. for each group (**P* < 0.05 by Mann–Whitney *U*-test).



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | PLVAP in liver is induced early during fetal organogenesis and is selectively expressed in sinusoidal endothelial cells. **a, b**, Immunohistological stains of formalin-fixed, paraffin-embedded fetal liver sections of wild-type (**a**) and *Plvap*^{-/-} (**b**) mice with anti-PLVAP antibody MECA-32 at the indicated time points. White arrows, representative vessels (brown). **c, d**, qPCR analyses of *Lyve1* expression in the liver in wild-type and *Plvap*^{-/-} mice (**c**), and of *Plvap* and *Lyve1* expression in sorted CD11b⁺F4/80^{intermediate} and CD11b⁺F4/80^{high} cells isolated from livers of E16.5 wild-type mice (cells pooled from 10 embryos). Liver denotes mRNA from whole E12.5 fetal liver (a positive control) (**d**). **e–g**, Immunofluorescent stains of E12.5, E14.5, and E16.5

fetal liver sections for PLVAP, F4/80 and LYVE-1 (**e**), E14.5 fetal liver for PLVAP and F4/80 (**f**), and E12.5, E14.5 and E16.5 liver for PLVAP and CD31 (another vascular endothelial marker) (**g**). Red arrow in **f** denotes an endothelial-penetrating protrusion of an F4/80⁺ myeloid cell. **h**, Immunofluorescent stains of B6;129 and BALB/c wild-type livers at E14.5 for PLVAP and LYVE-1. **i**, Immunofluorescent stains of human fetal liver (week 18) with an anti-PLVAP antibody (PAL-E). Shown are representative images ($n \geq 3$ (**a, b, e–h**) and $n = 1$ (**i**) biological replicates from 2 independent stains). Scale bars, 20 μm (**a, b, e–i**). In **c**, each dot represents one embryo (pooled from 2 independent experiments and from 2–3 litters, see Source Data), data are mean \pm s.e.m. of each group.

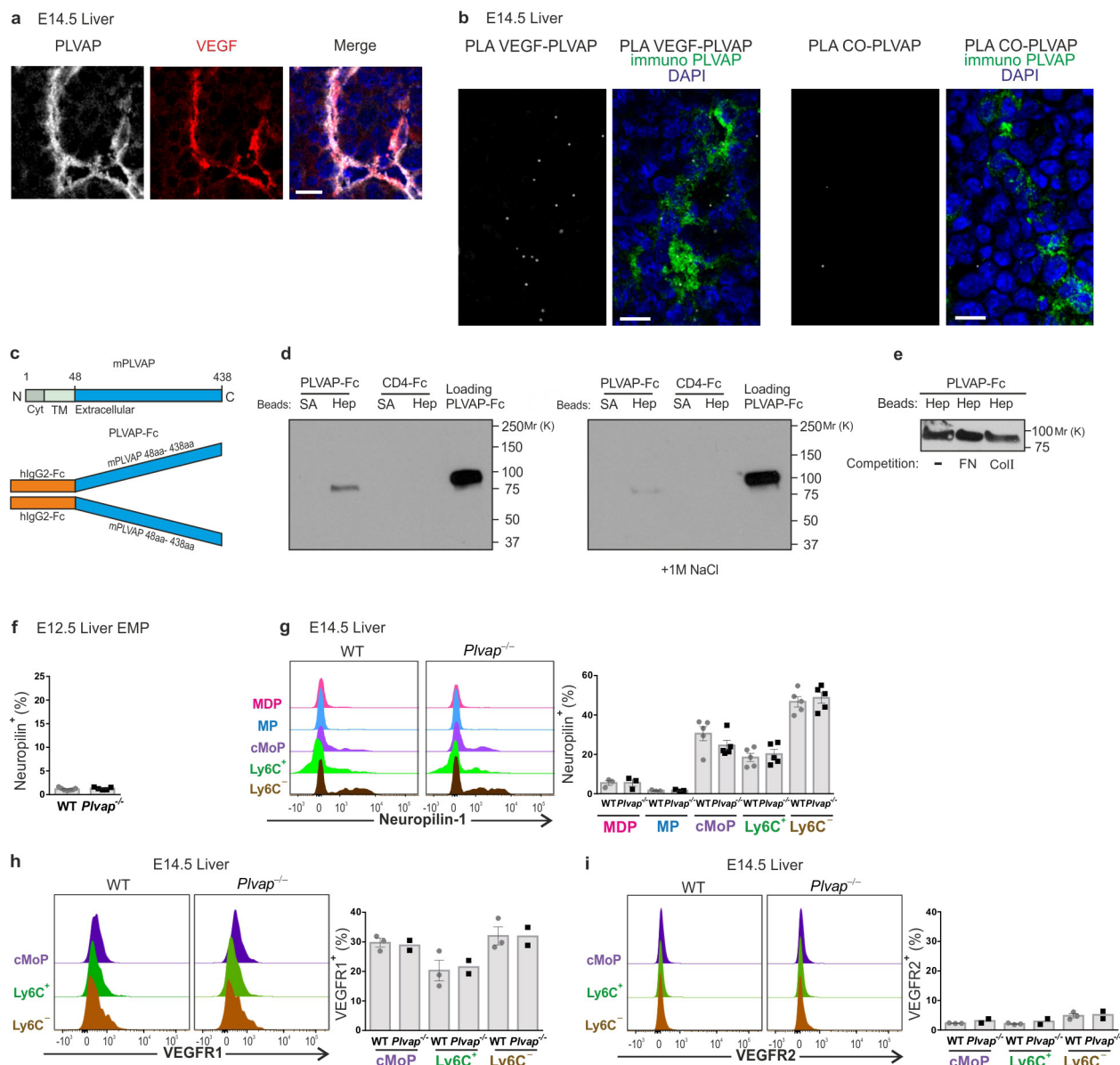


Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Time-selective and cell type-selective knock-down of PLVAP results in accumulation of fetal liver monocytes.

a, The timing of single tamoxifen injections and tissue collections with *Plvap*^{F/F}; *CAGGCre-ER*TM mice. **b**, Quantification of *Plvap* mRNA synthesis in fetal livers by qPCR 2 days after the tamoxifen treatments. **c**, Immunofluorescent stains of PLVAP and LYVE-1 in the fetal liver at E13.5, E14.5 and E15.5 (2 days after the tamoxifen pulse given on E11.5, E12.5 and E13.5, respectively). **d**, The gating strategy and flow cytometry analyses of the frequencies of fetal liver macrophage-dendritic cell precursors, myeloid precursors, common monocyte progenitors, Ly6C⁺ and Ly6C⁻ monocytes, (gates as in Extended Data Fig. 5c and Supplementary Table 1) at E13.5, E14.5 and E15.5 in *Plvap*^{F/F} (control) and *Plvap*^{F/F}; *CAGGCre-ER*TM mice (in each case 2 days after the tamoxifen

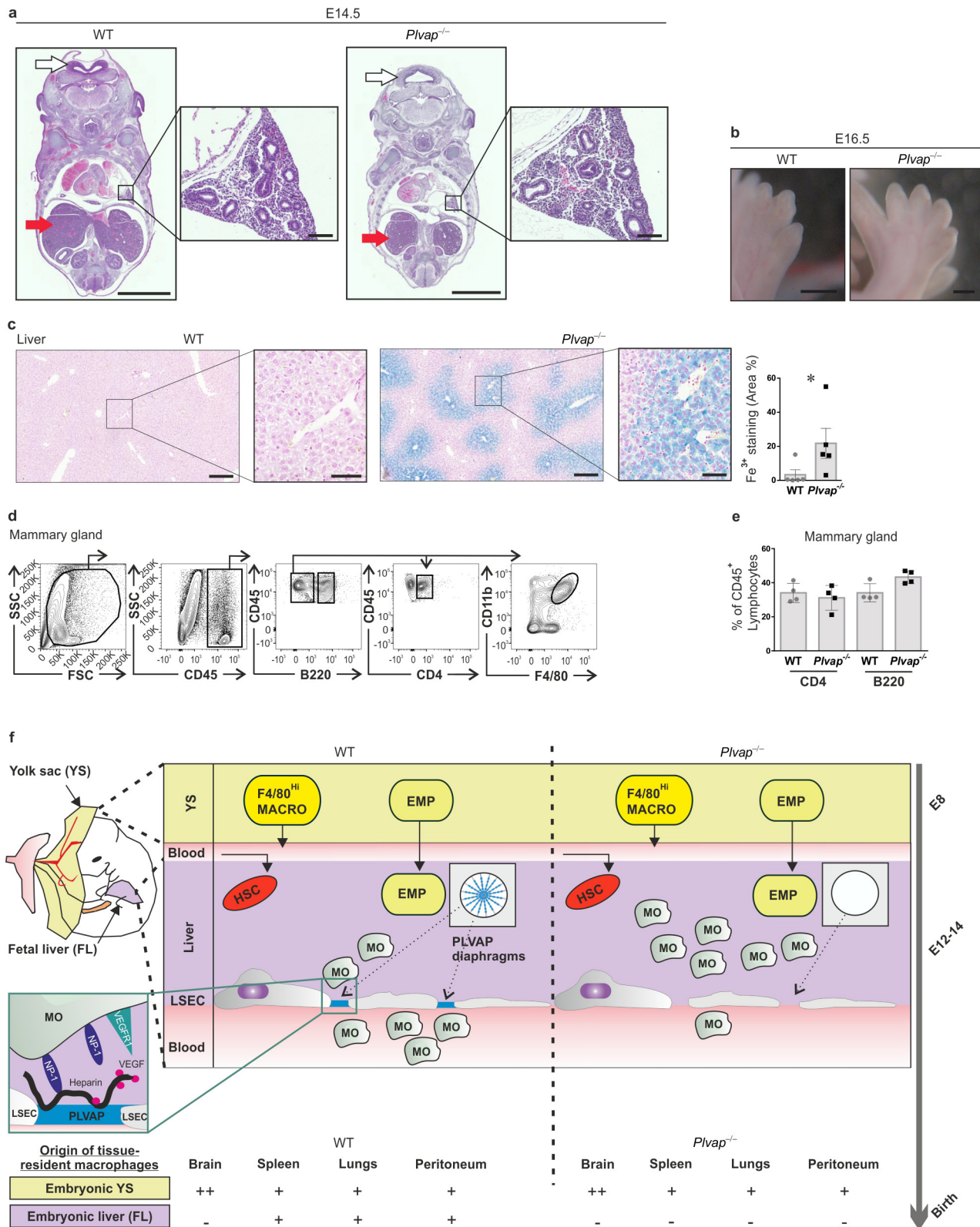
injection). **e**, Immunohistological analyses of PLVAP protein expression in the E14.5 fetal liver of *Plvap*^{F/F}; *Lyve1-Cre* and control (*Plvap*^{F/F}) mice. **f**, Flow cytometry analyses of macrophage-dendritic cell precursors, myeloid precursors, common monocyte progenitors, Ly6C⁺ and Ly6C⁻ monocytes (defined as in **d**) at E13.5 and E14.5 in the fetal liver of the control (*Plvap*^{F/F}) and *Plvap*^{F/F}; *Lyve1-Cre* mice. Shown are representative images ($n = 2$ (**c**) and $n = 4$ (**e**) biological replicates from 2 independent stains). Scale bars, 20 μm (**c**, **e**). The flow cytometry data are shown as frequency of live-gated CD45⁺ cells. Each dot represents one embryo (pooled from 2–3 independent experiments and from 2–4 litters (**b**, **d**, **f**), see the Source Data), data are mean \pm s.e.m. for each group (* $P < 0.05$ by Mann–Whitney U-test).



Extended Data Figure 9 | Interactions of PLVAP with heparin and VEGF, and VEGF receptor expression on fetal liver monocytes.

a, Immunofluorescent analysis of VEGF and PLVAP expression in E14.5 liver of wild-type mice ($n = 3$). Scale bar, 10 μ m. **b**, Proximity ligation assays (PLAs) in E14.5 fetal livers. Shown are PLA signals (small dots, white) between VEGF and PLVAP (leftmost) and between a negative control protein (CO) and PLVAP (third image from the left). In the merged images, total PLVAP expression determined by immunohistochemistry (immuno PLVAP, green) and nuclear stains (DAPI, blue) are displayed in addition to the PLA signals. Representative images from 3 independent experiments are shown. Scale bars, 10 μ m. **c**, Schematic depiction of mouse PLVAP protein and the PLVAP-Fc fusion protein. Cyt, cytoplasmic domain; TM, transmembrane domain. Numbers represent amino acids. **d**, Pull-down assays analysing the affinity of binding of PLVAP-Fc and CD4-Fc (a negative control) fusion proteins to heparin (Hep)- and streptavidin (Sa)-affinity (negative control) beads. The bead-bound proteins were eluted in Laemmli's sample buffer (the immunoblot on the left; this is the full image and different exposure of Fig. 3g), or in 1.0 M NaCl (the immunoblot on the right). The released proteins were separated in SDS-PAGE under reducing conditions and

visualized using immunoblotting for the Fc-tail. Aliquots of PLVAP-Fc (loading Plvap-Fc) were used as loading controls. Representative blots from two independent assays are shown. **e**, Pull-down assays analysing the binding of PLVAP-Fc (20 μ g) to heparin-beads in the absence (-) and presence of competing proteins fibronectin (FN, 100 μ g) and collagen (Coll, 100 μ g). The bound proteins were eluted, separated in SDS-PAGE under reducing conditions and visualized using immunoblotting for the Fc-tail. **f**, **g** Flow cytometry analyses of neuropilin-1 expression on live-gated CD41⁺CD45⁺c-Kit^{high}F4/80⁻ EMP cells in E12.5 livers (**f**), and on macrophage-dendritic cell precursors, myeloid precursors, common monocyte progenitors, Ly6C⁺ and Ly6C⁻ monocytes during monocytopoiesis in E14.5 livers (**g**) (gates as in Extended Data Fig. 5c and Supplementary Table 1) of wild-type and *Plvap*^{-/-} mice. **h**, **i**, Flow cytometry analyses of VEGFR1 (**h**) and VEGFR2 (**i**) expression on cMoP, Ly6C⁺ and Ly6C⁻ monocytes in E14.5 livers. The flow cytometry data are shown as frequency of live-gated (**f**) and live-gated CD45⁺ (**h**, **i**) cells. Each dot represents one embryo (pooled from 2–3 independent experiments and from 2–3 litters (**f**–**i**), see Source Data), data are mean \pm s.e.m. for each group.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Normal yolk-sac macrophage-dependent morphogenesis in *Plvap*^{-/-} mice and the function of PLVAP during macrophage ontogenesis.

a, Haematoxylin–eosin stains of coronal sections of whole E14.5 embryos from wild-type and *Plvap*^{-/-} mice. White arrows, neural tubes; red arrows, livers. Inset, the developing bronchial tree in the lungs. Scale bars, 2 mm (main image), 200 μ m (inset). **b**, Macroscopic images of toes in wild-type and *Plvap*^{-/-} embryos at E16.5. Scale bars, 1 mm. **c**, Prussian blue stains of livers, and quantification of Fe³⁺-containing cells (blue) in 5-week-old mice. Scale bars, 200 μ m (main image), 50 μ m (inset). **d**, The gating strategy for CD11b⁺F4/80⁺ cells in the mammary fat pad. **e**, Flow cytometry analyses of CD4⁺ T-helper cells and B220⁺ B lymphocytes in the mammary fat pads of wild-type and *Plvap*^{-/-} mice. Shown are representative images (**a–c**; $n = 3$ biological replicates). Flow cytometry data are shown as frequency of live-gated CD45⁺ leukocytes. Each dot represents one mouse embryo (pooled from 2 independent experiments, see the Source Data), data are mean \pm s.e.m. for each group. **f**, A schematic model depicting the organ-selective role of PLVAP in the seeding of fetal liver monocyte-derived tissue-resident macrophages. The yolk-sac-derived tissue-resident macrophages and

the progenitors (EMP and HSC) of fetal liver monocytes develop and seed normally in the absence of PLVAP. By contrast, PLVAP supports the egress of fetal liver-derived monocytes to the blood, and thereby the seeding of fetal liver-derived tissue-resident macrophages in different tissues. PLVAP fibrils form the fenestral diaphragms (blue) in fetal liver sinusoidal endothelial cells (LSEC). The seeding of bone-marrow-derived monocytes and macrophages after birth is PLVAP-independent (not shown). In the inset, the molecular interactions of PLVAP with heparin, neuropilin-1 (NP-1) and VEGF in the fetal liver and the expression of neuropilin-1 and VEGFR1 by E14.5 fetal liver monocytes (MO) are depicted. Although it remains to be experimentally tested, it is likely that PLVAP–heparin complexes at the diaphragms of fetal LSEC have the potential to regulate monocyte egress by providing an adhesive substrate for fetal liver monocytes (for example, via neuropilin-1, and probably via other molecules as well), and/or by immobilizing chemoattractants (for example, VEGF, and possibly other heparin-binding chemotactic molecules). The possible selective role of PLVAP in the vasculature of target organs during the seeding of fetal liver-derived monocytes also remains to be tested.

The epichaperome is an integrated chaperome network that facilitates tumour survival

Anna Rodina^{1*}, Tai Wang^{1*}, Pengrong Yan^{1*}, Erica DaGama Gomes^{1*}, Mark P. S. Dunphy^{2*}, Nagavarakishore Pillarsetty², John Koren III¹, John F. Gerecitano³, Tony Taldone¹, Hongliang Zong⁴, Eloisi Caldas-Lopes¹, Mary Alpaugh^{1†}, Adriana Corben⁵, Matthew Riolo¹, Brad Beattie⁶, Christina Pressl², Radu I. Peter⁷, Chao Xu¹, Robert Trondl¹, Hardik J. Patel¹, Fumiko Shimizu¹, Alexander Bolaender¹, Chenghua Yang¹, Palak Panchal¹, Mohammad F. Farooq⁸, Sarah Kishinevsky¹, Shanu Modi⁹, Oscar Lin⁵, Feixia Chu⁸, Sujata Patil¹⁰, Hediye Erdjument-Bromage¹¹, Pat Zanzonico⁶, Clifford Hudis⁹, Lorenz Studer¹², Gail J. Roboz⁴, Ethel Cesarman⁴, Leandro Cerchietti⁴, Ross Levine¹³, Ari Melnick⁴, Steven M. Larson², Jason S. Lewis², Monica L. Guzman^{4§} & Gabriela Chiosis^{1,9§}

Transient, multi-protein complexes are important facilitators of cellular functions. This includes the chaperome, an abundant protein family comprising chaperones, co-chaperones, adaptors, and folding enzymes—dynamic complexes of which regulate cellular homeostasis together with the protein degradation machinery^{1–6}. Numerous studies have addressed the role of chaperome members in isolation, yet little is known about their relationships regarding how they interact and function together in malignancy^{7–17}. As function is probably highly dependent on endogenous conditions found in native tumours, chaperomes have resisted investigation, mainly due to the limitations of methods needed to disrupt or engineer the cellular environment to facilitate analysis. Such limitations have led to a bottleneck in our understanding of chaperome-related disease biology and in the development of chaperome-targeted cancer treatment. Here we examined the chaperome complexes in a large set of tumour specimens. The methods used maintained the endogenous native state of tumours and we exploited this to investigate the molecular characteristics and composition of the chaperome in cancer, the molecular factors that drive chaperome networks to crosstalk in tumours, the distinguishing factors of the chaperome in tumours sensitive to pharmacologic inhibition, and the characteristics of tumours that may benefit from chaperome therapy. We find that under conditions of stress, such as malignant transformation fuelled by MYC, the chaperome becomes biochemically ‘rewired’ to form a network of stable, survival-facilitating, high-molecular-weight complexes. The chaperones heat shock protein 90 (HSP90) and heat shock cognate protein 70 (HSC70) are nucleating sites for these physically and functionally integrated complexes. The results indicate that these tightly integrated chaperome units, here termed the epichaperome, can function as a network to enhance cellular survival, irrespective of tissue of origin or genetic background. The epichaperome, present in over half of all cancers tested, has implications for diagnostics and also provides potential vulnerability as a target for drug intervention.

To investigate the chaperome in tumours we first analysed HSP90, the most abundant chaperome member in human cells^{1,2}. In cultured

non-transformed cells and in normal primary breast tissue (NPT, the normal tissue surrounding or adjacent to the corresponding primary tumour) (Fig. 1a, b), HSP90 focused primarily as a single species at the predicted isoelectric point (pI) of 4.9. However, cancer cell lines analysed by this method contained a complex mixture of HSP90 species spanning a pI range of 4.5 to 6; HSP90 α and HSP90 β isoforms were part of these complexes. Furthermore, although all cancer cell lines contained a number of HSP90 complexes with pI < 4.9, a subset was enriched in HSP90 complexes with the unusual pI of ≥ 5 , herein referred to as ‘type 1’ cells. We refer to cancer cell lines that contained mainly complexes with pI < 4.9 as ‘type 2’ cells. This distinction in HSP90 complexes was also evident in primary tumours (Fig. 1b). The total levels of HSP90 were essentially identical among all analysed samples, irrespective of whether they were type 1 or type 2 (Fig. 1a; see further analyses).

Under denaturing conditions, HSP90 in type 1 tumours focused mainly at the pI of ~ 4.9 (Fig. 1c). We therefore directed our attention on proteins interacting with HSP90 as the main instrument for pI change in type 1 tumours. HSP90 is known to interact with several co-chaperones including activator of HSP90 ATPase homologue 1 (AHA1, also known as AHSA1), cell division cycle 37 (CDC37), and HSP70–HSP90 organizing protein (HOP, also known as stress-inducible phosphoprotein 1 (STIP1)) which links HSP90 to the HSP70 machinery. Each of these co-chaperones has a distinct role. CDC37 facilitates activation of kinases, AHA1 augments HSP90 ATPase activity, and HSP70 and HOP participate in the chaperoning of proteins^{2–5,13}. We observed that cultured cells and primary tumours enriched in the high pI HSP90 species were also enriched in high-molecular-weight, multimeric forms of HSP90 and of other essential chaperome members (Fig. 1d and Extended Data Fig. 2c–e).

We found that PU-H71, an HSP90 inhibitor that binds to HSP90 more strongly when HSP90 is complexed with co-chaperones and onco-client proteins^{7,18,19}, also bound HSP90 more tightly in type 1 than in type 2 cells (Extended Data Fig. 3a–j). This was independent of chaperome expression or intracellular ATP levels (as PU-H71 is an ATP competitor) (Extended Data Fig. 4). At the molecular level, and unlike the anti-HSP90 antibody H9010, the small fraction

¹Program in Chemical Biology, Sloan Kettering Institute, New York, New York 10065, USA. ²Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.

³Lymphoma Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁴Haematology and Medical Oncology, Department of Medicine, Weill Cornell Medical College, New York, New York 10065, USA. ⁵Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁶Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁷Department of Mathematics, Technical University of Cluj-Napoca, Cluj-Napoca 400114, Romania. ⁸Molecular, Cellular & Biomedical Sciences, University of New Hampshire, Durham, New Hampshire 03824, USA. ⁹Breast Cancer Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ¹⁰Department of Epidemiology-Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ¹¹Microchemistry and Proteomics Core, Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ¹²Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ¹³Human Oncology and Pathogenesis Program, Sloan Kettering Institute, New York, New York 10065, USA. [†]Present address: Department of Biology and Department of Biomedical and Translational Sciences, Rowan University, Glassboro, New Jersey 08028, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

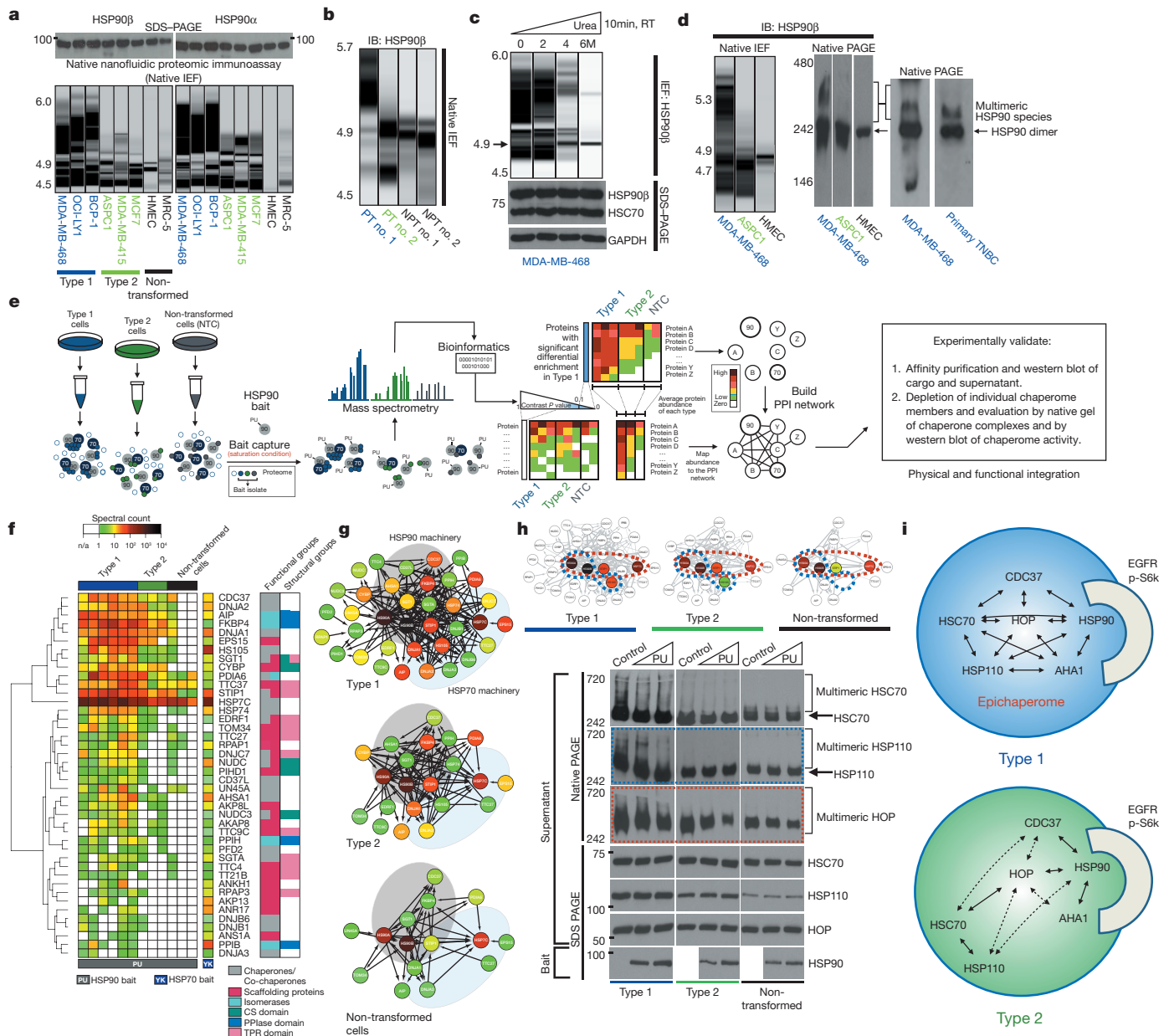


Figure 1 | A subset of cancer cells are enriched in stable multimeric chaperome complexes. **a–d**, The biochemical profile of indicated chaperome members in cell lines and primary specimens. IB, immunoblotting; TNBC, triple-negative breast cancer; NPT, the normal tissue surrounding or adjacent to the corresponding primary tumour; PT, primary tumour; RT, room temperature. The gel representation of the chromatogram is shown for IEF. See also Extended Data Fig. 2a, b. **e**, Workflow used to identify the chaperome components and establish their interconnectivity in cells. **f**, Heat map illustrating core HSP90 chaperome members enriched ($P < 0.1$) in type 1 tumours. Last lane, HSP70-interacting chaperome. **g**, Networks showing interactions

of cellular HSP90 that was part of the high-molecular-weight species enriched in type 1 tumours was most sensitive to PU-H71 (Extended Data Fig. 3c, h).

Our data suggested that a biochemically altered chaperome exists in type 1 tumours, so we investigated its composition (Fig. 1e–h). HSP90 protein isolates in type 1 tumours contained a significant enrichment of a number of chaperome proteins known to function as chaperone, co-chaperone, scaffolding, adaptor, interface mediators, foldase and isomerase proteins. Surprisingly, they also incorporated a large number of HSP70 chaperome regulators, in addition to the expected and known

between chaperome proteins. See also Extended Data Fig. 5. **h**, Changes in multimeric chaperome complexes and total chaperome levels in cell homogenates challenged with control or increasing concentrations of the HSP90-directed bait. All data were repeated independently twice with representative images shown. For uncropped gel data, see Supplementary Fig. 1. **i**, In both type 1 and 2 tumours, the HSP90 machinery is functional and regulates its onco-client proteins such as EGFR and p-S6K, but only type 1 but not type 2 tumours are characterized by stable, multimeric chaperome complexes that physically and functionally integrate the HSP90 and HSP70 machinery components.

HSP90 regulators (Supplementary Discussion). Similarly, an HSP70-directed bait isolated numerous HSP90 regulators in type 1 tumours (Fig. 1f and Extended Data Fig. 5a–e). Multiple connectivity networks that integrate the HSP90 and HSP70 machineries and expand their functional reach through participating scaffolding proteins were characteristic of type 1 but not of type 2 tumours or non-transformed cells (Fig. 1g). High-molecular-weight complexes that incorporate HSP90, HOP, HSC70 (heat shock cognate 70 kDa protein, the constitutively expressed HSP70 paralogue also known as HSPA8)² and its co-chaperone HSP110 (heat shock 105 kDa/110 kDa protein 1)², were present in

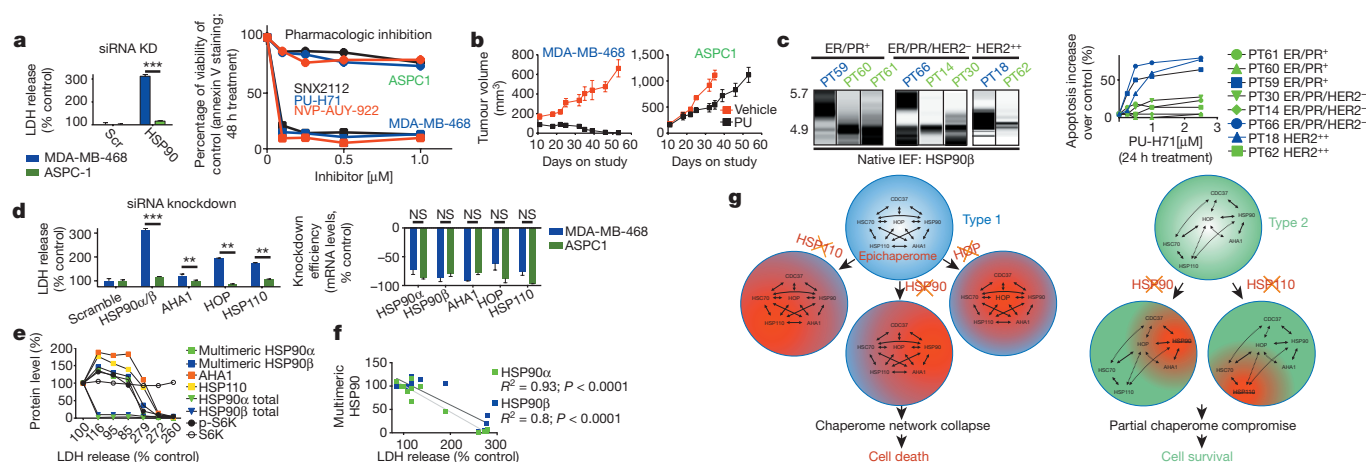


Figure 2 | The epichaperome facilitates cancer cell survival. **a**, Changes in cell viability upon HSP90 α and HSP90 β knockdown (mean \pm s.d., unpaired *t*-test, $n = 6$) or pharmacologic inhibition (mean, $n = 3$), as indicated. *** $P < 0.001$. **b**, Tumour volume in mice ($n = 5$) treated for the indicated time with PU-H71 or vehicle. Error bars show mean \pm s.d. **c**, Multimeric HSP90 complexes in primary breast cancer specimens ($n = 8$), clustered by biologic subtype, and their *ex vivo* sensitivity to PU-H71. **d**, Cytotoxicity upon siRNA knockdown of key chaperome members. For knockdown efficiency see messenger RNA (right)

type 1 but not in type 2 or non-transformed cells. The HSP90 bait readily depleted these multimeric species but left the non-bound HSC70 and HSP110 species unaltered (Fig. 1h). Similarly, dual knockdown of HSP90 α and HSP90 β or AHA1 modulated the high molecular HSC70 complexes only in type 1 tumours; knockdown of HSP110 modulated the multimeric HSP90 complexes only in type 1 tumours. Only the HOP knockdown modulated HSP90 and HSP70 complexes in both tumour types (Extended Data Fig. 5f, g). HSP90 was functional in both tumour types (chaperoned the kinases EGFR and p-S6K) and knockdown of HSP90 and AHA1 inhibited this activity in both tumour types (Extended Data Fig. 5g). Both the HSP90 α and HSP90 β paralogues, but mainly HSC70 and not HSP70 (the inducible HSP70 paralogue also known as HSP72 or HSP70-1)², participated in the reconfiguration of the chaperome in type 1 tumours (Extended Data Fig. 5h, i). Substantial reconfiguration of the chaperome organization only modestly affected the total chaperome levels (Extended Data Fig. 5g, h).

Together, these results lend support for the existence in type 1 tumours of HSP90- and HSP70-centric complexes that incorporate the

and protein levels (see Extended Data Fig. 5g). Error bars show mean \pm s.d., unpaired *t*-test, $n = 6$. ** $P < 0.01$; *** $P < 0.001$. LDH, lactate dehydrogenase. **e**, Epichaperome, total chaperome levels, chaperome activity and cell viability of type 1 cells in which several concentrations of siRNAs ($n = 7$) against HSP90 α and HSP90 β were titrated in. **f**, Correlative analysis between epichaperome levels and cell viability for data in **e** (Pearson's *r*, two-tailed, $n = 14$). See also Extended Data Fig. 7. For uncropped gel data, see Supplementary Fig. 1. **g**, Summary schematic.

co-chaperones of both machineries and integrate the chaperome into a large functional and physical network (Fig. 1h). Through scaffolding, adaptor, and interface modulator proteins, they bridge the chaperome to numerous cellular processes vital for tumour cell function. We refer to this highly integrated chaperome network of type 1 tumours as the epichaperome. Only a fraction of the entire chaperome pool participates in the chaperome rewiring of type 1 tumours. In contrast to the integrated epichaperome found in type 1 cells, no such integration is found in normal cells and type 2 tumours. In those cases, the HSP90 machinery only loosely interacts with the HSP70 machinery, mainly through the ubiquitous HSP90–HOP–HSP70 connection. In type 2 tumours, the two major chaperome machineries co-exist as insular chaperome communities that are only partially connected to each other.

Next, we investigated the functional relevance of the integrated epichaperome as compared to that of individual chaperome members or individual chaperome machineries. First, we investigated the reliance of type 1 and type 2 tumours on individual chaperome members (Fig. 2). In cells with essentially identical HSP90 levels, targeting of both HSP90 α

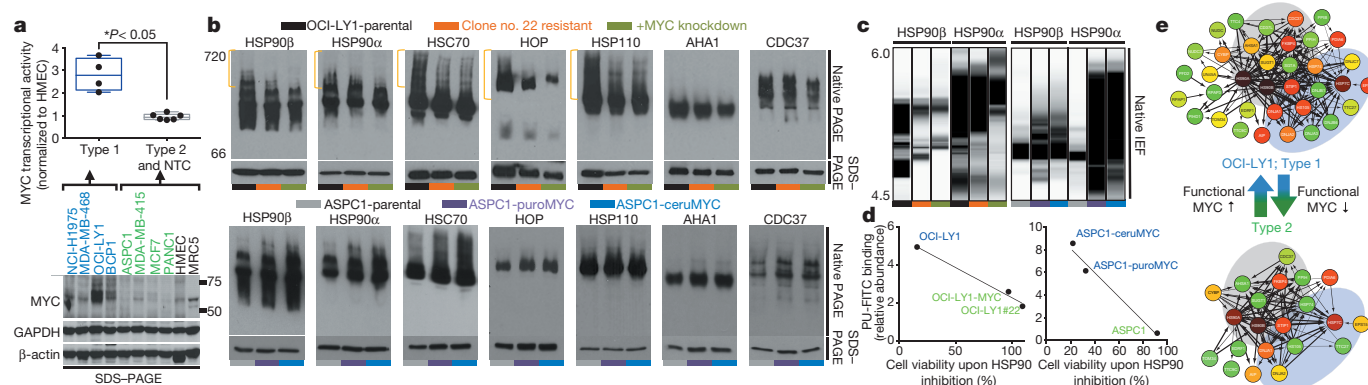


Figure 3 | MYC is a driver of chaperome rewiring into the epichaperome. **a**, MYC transcriptional activity (top) and protein levels (bottom) in the indicated cells. β -actin and GAPDH, loading controls (unpaired *t*-test, each data point is the mean of two technical replicates and represents a cell line). HMEC, human mammary epithelial cells. **b–d**, Changes in multimeric chaperome complexes (top) and total protein

(bottom) in the indicated homogenates of cells where MYC levels were modulated as indicated. Data were repeated independently twice with representative data shown. In **d** each data point is the mean from two independent experiments and represents a cell line. For uncropped gel data, see Supplementary Fig. 1. **e**, Summary schematic showing MYC as the cellular switch for epichaperome assembly and disassembly.

and HSP90 β was toxic to type 1 but not type 2 tumours (Fig. 2a). We confirmed this in mice bearing xenograft tumours (Fig. 2b), in primary specimens *ex vivo* (Fig. 2c), and in human patients (Extended Data Fig. 6a). By contrast, while being toxic only to type 1 tumours, targeting of HSP90 inactivated its chaperone activity in both tumour types (see inhibition and/or degradation of HSP90 regulated proteins and pathways, such as EGFR and PI3K/AKT, and cell growth inhibition with similar half-maximum inhibitory concentration (IC₅₀) potencies; Extended Data Fig. 6b–i). As observed for HSP90, downregulation of other individual chaperome members led to cell death in type 1 but not in type 2 cells (Fig. 2d). When we interfered with epichaperome formation by reducing the levels of one of its components, AHA1, cells became less amenable to killing by PU-H71 (Extended Data Fig. 7a). From these observations, we propose that the epichaperome has a role as the survival facilitator of type 1 tumours. In type 1 tumours, we observed that a striking decline (>95% at protein level, Fig. 2e and Extended Data Fig. 7b–f) in overall total HSP90 levels was initially paralleled by an increase in the epichaperome. This occurred by an increased production of chaperome members that presumably sequestered the remaining HSP90 into the high-molecular-weight complexes. Under these conditions of low total HSP90 but high epichaperome levels, no cell death was observed. HSP90 function was also not impaired (see EGFR and p-S6K). As HSP90 levels continued to drop, however, an inflection was observed. At this point, the epichaperome levels dropped, chaperome members were depleted, a sudden drop in HSP90 function occurred, and cell death ensued. Epichaperome expression significantly correlated with cell viability (Fig. 2f). In contrast, in type 2 cells, a similar drastic reduction of HSP90 levels halted its activity, but failed to re-wire the chaperome into the epichaperome and did not result in cell death (Extended Data Fig. 7g). Together, these findings are consistent with the formation of the epichaperome as a survival mechanism for type 1 tumours. When the epichaperome is dismantled by ablation of a chaperome component, the network collapses and leads to cell death. In type 2 tumours in which the integration of the chaperome is only partial and most chaperome members function as insular communities, depletion of a chaperome member only ‘locally’ compromises the chaperome, while overall cellular survival is maintained (Fig. 2g).

To understand the molecular mechanisms leading to epichaperome formation, we investigated the HSP90 interactome in type 1 and type 2 cells (Extended Data Fig. 5b). We identified MYC as a transcription factor that could most probably explain the protein signature observed in type 1 tumours. MYC target genes were significantly enriched in type 1 tumours, as was a MYC transcriptional signature and positive

regulators of MYC expression/function (Extended Data Fig. 8). We experimentally confirmed a significantly higher MYC transcriptional activity in type 1 versus type 2 cancer cells (Fig. 3a). Knockdown of MYC re-wired type 1 cells into type 2; this was reflected in the composition of the chaperome complexes, the binding to PU-H71 and reduced sensitivity to HSP90 inhibition (Fig. 3b–d). We also observed a decrease in MYC mRNA and protein levels in type 1 cells that became resistant to HSP90 inhibitors after long-term treatment with suboptimal HSP90 inhibitor concentrations, and demonstrated their rewiring into type 2 cells (Fig. 3b–d and Extended Data Fig. 9a–e). However, the introduction of a functional MYC gene into a type 2 cancer cell rewired it to become type 1 (Fig. 3b–d and Extended Data Fig. 9f–l). Oncogenes that require single chaperome machinery activity for sufficient transformative power (for example, vSRC and mutated MET kinase require HSP90 (ref. 3)) were unable to induce the formation of the epichaperome, nor was epichaperome formation necessary to buffer their oncogenic function (Extended Data Fig. 9m–p). Together these findings suggest that the transcription factor, MYC, at least in part, causes the molecular rewiring of the chaperome into the epichaperome as observed in type 1 tumours (Fig. 3e).

The data point to the epichaperome and not the individual chaperome members as potential targets of chaperome-directed intervention in cancer. We thus assessed the prevalence of tumours expressing the epichaperome complexes. In probing cancer cell lines representing pancreatic, gastric, lung, and breast cancers, as well as lymphomas and leukaemias, we found that approximately 60–70% presented medium to high levels of epichaperome complexes (Fig. 4a). Similar results were obtained with primary liquid tumours (Fig. 4b) and solid tumours including lymphomas (Fig. 4c). This establishes that over half of tumours tested use the epichaperome irrespective of their subtype, provenance or genetic background. Toxicity to HSP90 inhibition correlated with the presence of the epichaperome ($P = 0.0006$; $R^2 = 0.71$) but was independent of the levels of chaperome members, HSP90 client proteins, anti-apoptotic proteins and genetic alterations. This correlation held in over 90 cell lines encompassing breast cancer, lung cancer, pancreatic and gastric cancers, leukaemia and lymphomas ($P < 0.0001$; $R^2 = 0.44$) and was true for several HSP90 inhibitors^{19,20} (Extended Data Fig. 10). The data thus indicate that it is the abundance of the epichaperome, and not merely its existence, that is indicative of the reliance of tumours on the epichaperome. If patients were to be selected for epichaperome therapy, not only the existence of this species but also its abundance should be measured. To further confirm these findings, we collected primary breast cancer specimens ($n = 40$)

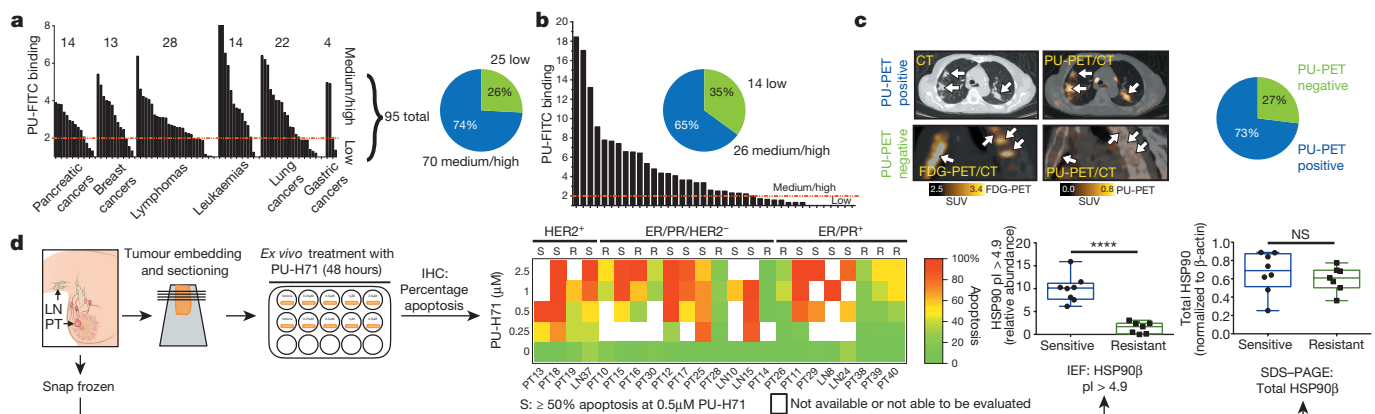


Figure 4 | More than half of all tumours tested express the epichaperome. **a–c**, Epichaperome measurement (abundance measured by PU-FITC, see Methods) in a panel of 95 cancer cell lines (**a**), 40 primary acute myeloid leukaemias (AMLs) (**b**), and epichaperome detection (by PU-PET, see Methods) in 51 solid tumours and lymphomas, in patients (**c**). Each bar represents a cell line; data are the mean from two independent experiments. For PU-PET, cross-sectional CT and PU-PET

images of representative tumours are shown each at the same transaxial plane. Location of the tumours is indicated by arrows. Scale bars, PET window display intensity scales, with upper and lower standardized uptake value (SUV) thresholds. **d**, *Ex vivo* apoptotic sensitivity of primary breast tumours to PU-H71 ($n = 23$) was compared to epichaperome abundance and total HSP90 levels. PT, primary tumour; LN, lymph node (error bars represent mean \pm s.d., unpaired *t*-test, $n = 15$). **** $P < 0.0001$.

obtained from surgery. Of these, 23 specimens were suitable samples to be evaluated for PU-H71 sensitivity and/or epichaperome abundance by isoelectric focusing (IEF) (that is, HSP90 of a pI > 4.9) and total HSP90 by SDS-PAGE (Fig. 4d). In these specimens, we found a spectrum of sensitivities, ranging from 0% to 100% for apoptotic response, with 56% undergoing at least 50% apoptosis when challenged with 0.5 μ M PU-H71. Abundance of the epichaperome but not of total HSP90 significantly correlated with sensitivity (Fig. 4d, $P < 0.0001$).

Here we report the discovery of a new mechanism of tumour regulation. Our study unveils a novel usage of the chaperome in epichaperome networks for cancer cell survival. The epichaperome results from changes in the chaperome that are driven by a change in the cellular milieu, that is, activation of MYC, rather than defects in the chaperome members, composition, number or structure. It manifests as an enhanced physical integration of the HSP90 and HSP70 machineries, resulting in the utilization of their capacities in the tumour cell environment, and thereby also presenting a vulnerability that might possibly be exploited therapeutically with pharmacological modulators. Our results offer a blueprint for the future development of therapeutic inhibitors specific for multimeric chaperome complexes, and might encourage further drug developments and advances in innovative companion diagnostics (Extended Data Fig. 1).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 March; accepted 2 September 2016.

Published online 5 October 2016.

- Brehme, M. *et al.* A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Reports* **9**, 1135–1150 (2014).
- Finka, A. & Goloubinoff, P. Proteomic data from human cell cultures refine mechanisms of chaperone-mediated protein homeostasis. *Cell Stress Chaperones* **18**, 591–605 (2013).
- Taipale, M. *et al.* A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* **158**, 434–448 (2014).
- Echtenkamp, F. J. & Freeman, B. C. Expanding the cellular molecular chaperone network through the ubiquitous cochaperones. *Biochim. Biophys. Acta* **1823**, 668–673 (2012).
- Echeverría, P. C., Bernthaler, A., Dupuis, P., Mayer, B. & Picard, D. An interaction network predicted from public data as a discovery tool: application to the Hsp90 molecular chaperone machine. *PLoS One* **6**, e26044 (2011).
- Balchin, D., Hayer-Hartl, M. & Hartl, F. U. *In vivo* aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).
- Moulick, K. *et al.* Affinity-based proteomics reveal cancer-specific networks coordinated by Hsp90. *Nature Chem. Biol.* **7**, 818–826 (2011).
- Woodford, M. R. *et al.* Impact of posttranslational modifications on the anticancer activity of Hsp90 inhibitors. *Adv. Cancer Res.* **129**, 31–50 (2016).
- Kamal, A. *et al.* A high-affinity conformation of Hsp90 confers tumour selectivity on Hsp90 inhibitors. *Nature* **425**, 407–410 (2003).
- Alarcon, S. V. *et al.* Tumor-intrinsic and tumor-extrinsic factors impacting Hsp90-targeted therapy. *Curr. Mol. Med.* **12**, 1125–1141 (2012).
- Mollapour, M. & Neckers, L. Post-translational modifications of Hsp90 and their contributions to chaperone regulation. *Biochim. Biophys. Acta* **1823**, 648–655 (2012).
- Assimon, V. A., Southworth, D. R. & Gestwicki, J. E. Specific binding of tetratricopeptide repeat proteins to heat shock protein 70 (Hsp70) and heat shock protein 90 (Hsp90) is regulated by affinity and phosphorylation. *Biochemistry* **54**, 7120–7131 (2015).
- Zuehlke, A. & Johnson, J. L. Hsp90 and co-chaperones twist the functions of diverse client proteins. *Biopolymers* **93**, 211–217 (2010).
- Shrestha, L. & Young, J. C. Function and chemotypes of human Hsp70 chaperones. *Curr. Top. Med. Chem.* **16**, 2812–2828 (2016).
- Smith, J. R. *et al.* Restricting direct interaction of CDC37 with HSP90 does not compromise chaperoning of client proteins. *Oncogene* **34**, 15–26 (2015).
- Radons, J. The human HSP70 family of chaperones: where do we stand? *Cell Stress Chaperones* **21**, 379–404 (2016).
- Patel, P. D. *et al.* Paralog-selective Hsp90 inhibitors define tumor-specific regulation of HER2. *Nature Chem. Biol.* **9**, 677–684 (2013).
- Nayar, U. *et al.* Targeting the Hsp90-associated viral oncoproteome in gammaherpesvirus-associated malignancies. *Blood* **122**, 2837–2847 (2013).
- Shrestha, L., Patel, H. J. & Chiosis, G. Chemical tools to investigate mechanisms associated with HSP90 and HSP70 in disease. *Cell Chemical Biology* **23**, 158–172 (2016).
- Taldone, T., Ochiana, S. O., Patel, P. D. & Chiosis, G. Selective targeting of the stress chaperome as a therapeutic strategy. *Trends Pharmacol. Sci.* **35**, 592–603 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Gewirth, A. Younes, J. Young and E. Devaney for reading the manuscript. This work was supported by the NIH (R01 CA172546, R01 CA155226, P01 CA186866, P30 CA08748, P50 CA86438, DP2 OD007399), the Jane H. Gordon Breast Cancer Research Fund, the Breast Cancer Research Fund, Susan G. Komen for the Cure, W. H. Goodwin, A. Goodwin and the Commonwealth Foundation for Cancer Research, and the Experimental Therapeutics Center of MSKCC, the Hirshberg Foundation for Pancreatic Cancer, the Irma T. Hirsch Foundation and the Rubenstein Center for Pancreatic Research. A.R. was supported by the National Center for Advancing Translational Sciences of the NIH under award number UL1TR000457.

Author Contributions A.R., T.W., P.Y., E.D.G., M.P.S.D., M.R., J.F.G., M.A., N.P., A.C., H.Z., J.K. III, B.B., C.P., C.X., R.T., H.J.P., P.P., M.F.F., F.S., E.C.-L., A.B., S.K., C.Y., O.L. and H.E.-B. performed experiments. S.M., T.T., C.H., N.P., L.C., A.M., E.C., L.S., J.S.L. and G.J.R. provided reagents. N.P., A.R., E.D.G., T.W., P.Y., P.Z., B.B., R.L., A.M., S.M.L., J.S.L., M.P.S.D., R.I.P., F.C., J.F.G., H.E.-B., G.C. and M.L.G. participated in the design and analysis of various experiments. S.P., R.I.P. and T.W. performed statistical analyses and G.C. and M.L.G. wrote the paper. All authors read the manuscript and/or discussed the experiments.

Author Information Primary data, such as raw mass spectrometry files, Mascot generic format files and proteomics data files created by Scaffold have been deposited in the MassIVE database (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>); MassIVE accession ID: MSV000079877. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.C. (chiosisg@mskcc.org) or M.L.G. (mlg2007@med.cornell.edu).

Reviewer Information Nature thanks J. W. Kelly and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Reagents. HSP90 inhibitors used in this study including PU-H71, PU-DZ13, NVP-AUY922, and SNX-2112 were synthesized as previously reported^{7,19}. 17-DMAG was purchased from Sigma. HSP90 bait (PU-H71 beads)²¹, HSP70 bait (YK beads)²², biotinylated YK (YK-biotin)²², fluorescently labelled PU-H71 (PU-FITC)²³, the control derivatives PU-TEG and PU-FITC9 (ref. 24), and the radiolabelled PU-H71-derivative ¹²⁴I-PU-H71 (ref. 25) were generated as previously described. The specificity of PU-H71 for HSP90 and over other proteins was extensively analysed⁷. Thus binding of PU-H71 in cell homogenates, live cells and organisms denotes binding to HSP90 species characteristic of each analysed tumour or tissue. Combined with the findings that PU-H71 binds more tightly to HSP90 in type 1 than in type 2 cells, an observation true for cell homogenates, live cells, and *in vivo*, at the organismal level, we propose that labelled versions of PU-H71 are reliable tools to perturb, identify and measure the expression of the high-molecular-weight, multimeric HSP90 complexes in tumours. The specificity of YK probes for HSP70 was previously reported^{22,26–28}.

Cell lines. Cell lines were obtained from laboratories at WCMC or MSKCC, or were purchased from the American Type Culture Collection (ATCC) or Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). Cells were cultured as per the providers' recommended culture conditions. Cells were authenticated using short tandem repeat profiling and tested for mycoplasma. The pancreatic cancer cell lines include: ASPC-1 (CRL-1682), PL45 (CRL-2558), MiaPaCa2 (CRL-1420), SU.86.86 (CRL-1837), CFPAC (CRL-1918), Capan-2 (HTB-80), BxPc-3 (CRL-1687), HPAFII (CRL-1997), Capan-1 (HTB-79), Panc-1 (CRL-1469), Panc05.04 (CRL-2557) and Hs766t (HTB-134) (purchased from the ATCC); 931102 and 931019 are patient derived cell lines provided by Y. Janjigian, MSKCC. Breast cancer cell lines were obtained from ATCC and include MDA-MB-468 (HTB-132), HCC1806 (CRL-2335), MDA-MB-231 (CRM-HTB-26), MDA-MB-415 (HTB-128), MCF-7 (HTB-22), BT-474 (HTB-20), BT-20 (HTB-19), MDA-MB-361 (HTB-27), SK-Br-3 (HTB-30), MDA-MB-453 (HTB-131), T-47D (HTB-133), AU565 (CRL-2351), ZR-75-30 (CRL-1504), ZR-75-1 (CRL-1500). Lymphoma cell lines include: Akata1, Mutu-1 and Rae-1 (provided by W. Tam, WCMC); BCP-1 (CRL-2294), Daudi (CCL-213), EB1 (HTB-60), NAMALWA (CRL-1432), P3HR-1 (HTB-62), SU-DHL-6 (CRL-2959), Farage (CRL-2630), Toledo (CRL-2631) and Pfeiffer (CRL-2632) (obtained from ATCC); HBL-1, MD901 and U2932 (kindly provided by J. Angel Martinez-Climent, Centre for Applied Medical Research, Pamplona, Spain); Karpas422 (ACC-32), RCK8 (ACC-561) and SU-DHL-4 (ACC-495) (obtained from the DSMZ); OCI-LY1, OCI-LY3, OCI-LY4, OCI-LY7 and OCI-LY10 (obtained from the Ontario Cancer Institute); TMD8 (kindly provided by L. M. Staudt, NIH); BC-1 (derived from an AIDS-related primary effusion lymphoma); IBL-1 and IBL-4 (derived from an AIDS-related immunoblastic lymphoma) and BC3 (derived from a non-HIV primary effusion lymphoma). Leukaemia cell lines include: REH (CRL-8286), HL-60 (CCL-240), KASUMI-1 (CRL-2724), KASUMI-4 (CRL-2726), TF-1 (CRL-2003), KG-1 (CCL-246), K562 (CCL-243), TUR (CRL-2367), THP-1 (TIB-202), U937 (CRL-1593.2), MV4-11 (CRL-9591) (obtained from ATCC); KCL-22 (ACC-519), OCI-AML3 (ACC-582) and MOLM-13 (ACC-554) (obtained from DSMZ). The lung cancer cell lines include: NCI-H3122, NCI-H299 (provided by M. Moore, MSKCC); EBC1 (provided by Dr Mellinshoff, MSKCC); PC9 (kindly provided by D. Scheinberg, MSKCC), HCC15 (ACC-496) (DSMZ), HCC827 (CRL-2868), NCI-H2228 (CRL-5935), NCI-H1395 (CRL-5868), NCI-H1975 (CRL-5908), NCI-H1437 (CRL-5872), NCI-H1838 (CRL-5899), NCI-H1373 (CRL-5866), NCI-H526 (CRL-5811), SK-MES-1 (HTB-58), A549 (CCL-185), NCI-H647 (CRL-5834), Calu-6 (HTB-56), NCI-H522 (CRL-5810), NCI-H1299 (CRL-5803), NCI-H1666 (CRL-5885) and NCI-H1703 (CRL-5889) (obtained from ATCC). The gastric cancer cell lines include: MKN74 (obtained from G. Schwarz, Columbia University), SNU-1 (CRL-5971) and NCI-N87 (CRL-5822) (obtained from ATCC), OE19 (ACC-700) (DSMZ). The non-transformed cell lines MRC-5 (CCL-171), human lung fibroblast and HMEC (PCS-600-010), human mammary epithelial cells were obtained from ATCC. NIH-3T3, and NIH-3T3 cell lines stably expressing either mutant MET (Y1248H) or vSRC, were provided by L. Neckers, National Cancer Institute (NCI), USA, and were previously reported^{29,30}.

Primary breast cancer specimens. Patient tissue was obtained with informed consent and authorized through institutional review board (IRB)-approved bio-specimen protocol number 09-121 at Memorial Sloan Kettering Cancer Centre (New York, New York). Specimens were treated for 24 h or 48 h with the indicated concentrations of PU-H71 as previously described³¹. Following treatment, slices were fixed in 4% formalin solution for 1 h, then stored in 70% ethanol. For tissue analysis, slices were embedded in paraffin, sectioned, slide-mounted, and stained with haematoxylin and eosin (H&E). Apoptosis and necrosis of the tumour cells (as percentage) was assessed by reviewing all the H&E slides of the case (controls and treated ones) *in toto*, blindly, allowing for better estimation of the overall treatment effect to the tumour. In addition, any effects to precursor lesions (if present) and

any off-target effects to benign surrounding tissue, were analysed. Tissue slides were assessed blindly by a breast cancer pathologist who determined the apoptotic events in the tumour, as well as any effect on adjacent normal tissue³¹.

Primary acute myeloid leukaemia (AML). Cryopreserved primary AML samples were obtained with informed consent and Weill Cornell Medical College IRB approval (IRB number 0910010677 and IRB number 0909010629). Samples were thawed and cultured for *in vitro* treatment as described previously³².

Clinical trials. The microdose ¹²⁴I-PU-H71 PET-CT (Dunphy, M. PET imaging of cancer patients using ¹²⁴I-PUH71: a pilot study available from: <http://clinicaltrials.gov>; NCT01269593) and phase I PU-H71 therapeutic (Gerecitano, J. The first-in-human phase I trial of PU-H71 in patients with advanced malignancies available from: <http://clinicaltrials.gov>; NCT01393509) studies were approved by the institutional review board (protocols 10-139 and 11-041, respectively), and conducted under an exploratory investigational new drug (IND) application approved by the US Food and Drug Administration. Patients provided signed informed consent before participation. ¹²⁴I-PU-H71 tracer was synthesized in-house by the institutional cyclotron core facility at high specific activity.

Epichaperome detection by PU-PET (¹²⁴I-PU-H71 positron emission tomography). For PU-PET, research PET-CT was performed using an integrated PET-CT scanner (Discovery DSTE, General Electric). CT scans for attenuation correction and anatomic coregistration were performed before tracer injection. Patients received 185 megabecquerel (MBq) of ¹²⁴I-PU-H71 by peripheral vein over two minutes. PET data were reconstructed using a standard ordered subset expected maximization iterative algorithm. Emission data were corrected for scatter, attenuation, and decay. ¹²⁴I-PU-H71 scans (PU-PET) were performed at 24 h after tracer administration. Each picture shown in Fig. 4c and Extended Fig. 6a is a scan taken of an individual patient. PET window display intensity scales for FDG and PU-PET fusion PET-CT images are given for both PU-PET and FDG-PET. Numbers in the scale bar indicate upper and lower SUV thresholds that define pixel intensity on PET images. The phase I trial included patients with solid tumours and lymphomas who had undergone prior treatment and currently had no curative treatment options. Patient cohorts were treated with PU-H71 at escalating dose levels determined by a modified continuous reassessment model. Each patient was treated with his or her assigned dose of PU-H71 on day 1, 4, 8, and 11 of each 21-day cycle.

Neurons. Human embryonic stem cells (hESCs) were differentiated with a modified dual-SMAD inhibition protocol towards floor plate-based midbrain dopaminergic (mDA) neurons as described previously³³. hESCs were maintained on mouse embryonic fibroblasts and passaged with Dispase (STEMCELL Technologies). For each differentiation, hESCs were harvested with Accutase (Innovative Cell Technology). At day 30 of differentiation, hESC-derived mDA neurons were replated and maintained on dishes precoated with polyornithine (PO; 15 µg ml⁻¹), laminin (1 µg ml⁻¹), and fibronectin (2 µg ml⁻¹) in Neurobasal/B27/L-glutamine-containing medium (NB/B27; Life Technologies) supplemented with 10 µM Y-27632 (until day 32) and with BDNF (brain-derived neurotrophic factor, 20 ng ml⁻¹; R&D), ascorbic acid (AA; 0.2 mM, Sigma), GDNF (glial cell line-derived neurotrophic factor, 20 ng ml⁻¹; R&D), TGFβ3 (transforming growth factor type β3, 1 ng ml⁻¹; R&D), dibutyryl cAMP (0.5 mM; Sigma), and DAPT (10 nM; Tocris). Two days after replating, mDA neurons were treated with 1 µg ml⁻¹ mitomycin C (Tocris) for 1 h to kill any remaining non-post mitotic contaminants. Assays were performed at day 65 of neuron differentiation.

Epichaperome abundance measurement using the PU-FITC flow cytometry assay. The PU-FITC assay was performed as previously described^{7,23}. Briefly, cells were incubated with 1 µM PU-FITC at 37 °C for 4 h. Then cells were washed twice with FACS buffer (PBS/0.5% FBS), and resuspended in FACS buffer containing 1 µg ml⁻¹ DAPI. HL-60 cells were used as internal control to calculate fold binding for all cell lines tested. The mean fluorescence intensity (MFI) of PU-FITC in treated viable cells (DAPI negative) was evaluated by flow cytometry. For primary AML specimens, cells were also stained with anti-CD45-APC-H7, to identify blasts and lymphocyte populations (BD biosciences). Blasts and lymphocyte populations were gated based on SSC versus CD45. The fold PU-FITC binding of leukaemic blasts (CD45dim) was calculated relative to lymphocytes (CD45hiSSCslow). The FITC derivative FITC9 was used as a negative control.

PU-FITC microscopy. Cells were seeded on coverslips in 6-well plate and cultured overnight. Cells were treated with 1 µM PU-FITC or negative control (PU-FITC9, an HSP90 inert PU-H71 derivative labelled with FITC). At 4 h post-treatment, cells were fixed with 4% formaldehyde at room temperature for 30 min, and the coverslips were mounted on slides with DAPI-Fluoromount-G Mounting Media (Southern Biotech). The images were captured using EVOS FL Auto imaging system (ThermoFisher Scientific) or a confocal microscope (Zeiss LSM5).

HSP90 immunofluorescence staining. Cells were seeded on coverslips and cultured overnight. Cells were fixed with 4% formaldehyde at room temperature for 30 min, washed three times with PBS, and permeabilized with 0.2% Triton X-100 in

blocking buffer (PBS/5% BSA) for 10 min. Cells were incubated in blocking buffer for 30 min, and then incubated with rabbit anti-human HSP90 α antibody (1:500, Abcam 2928) and mouse anti-human HSP90 β (1:500, Stressmarq H9010), or rabbit and mouse normal IgG, in blocking buffer for 1 h. Cells were washed three times with PBS, and incubated with goat anti-mouse Alexa Fluor 568 and goat anti-rabbit Alexa Fluor 488 (1:1,000, ThermoFisher Scientific) in blocking buffer in the dark for 1 h. Cells were then washed three times with PBS, and the coverslips were removed from the plate, and mounted on slides with DAPI-Fluoromount-G Mounting Media (Southern Biotech). The images were captured using EVOS FL Auto imaging system (ThermoFisher Scientific) or a confocal microscope (Zeiss LSM5). Fluorescence intensity was quantified by the integrated density algorithm as implemented in ImageJ.

PU-FITC or GM-cy3B binding to HSP90 in cell homogenates. Assays were carried out in black 96-well microplates (Greiner Microtron Fluotrac 200). A stock of 10 μ M PU-FITC (or GM-cy3B³⁴) was prepared in DMSO and diluted with Felts buffer (20 mM Hepes (K), pH 7.3, 50 mM KCl, 2 mM DTT, 5 mM MgCl₂, 20 mM Na₂MoO₄, and 0.01% NP40 with 0.1 mg ml⁻¹ BGG). To each well was added the fluorescent dye-labelled HSP90 ligand (3 nM PU-FITC or 6 nM GM-cy3B), and cell lysates (7.5 μ g) in a final volume of 100 μ l Felts buffer. For each assay, background wells (buffer only), and tracer controls (PU-FITC only) were included on assay plate. To determine the equilibrium binding of GM-cy3b, increasing amounts of lysate (up to 20 μ g of total protein) were incubated with tracer. The assay plate was placed on a shaker at room temperature for 60 min and the FP values in mP were measured every 5 min. At time $t = 60$ min, dissociation of fluorescent ligand was initiated by adding 1 μ M PU-H71 in Felts buffer to each well and then placing the assay plate on a shaker at room temperature and measuring the FP values in mP every 5 min. The assay window was calculated as the difference between the FP value recorded for the bound fluorescent tracer and the FP value recorded for the free fluorescent tracer (defined as mP – mPf). Measurements were performed on a Molecular Devices SpectraMax Paradigm instrument (Molecular Devices, Sunnyvale, CA), and data were imported into SoftMaxPro6 and analysed in GraphPad Prism 5.

Protein analysis by the NanoPro capillary-based immunoassay platform. To identify and separate chaperome complexes in tumours, and to overcome the limitations of classical protein chromatography methods for resolving complexes of similar composition and size, we took advantage of a capillary-based platform that combines isoelectric focusing (IEF) with immunoblotting capabilities³⁵. This methodology uses an immobilized pH gradient to separate native multimeric protein complexes based on their isoelectric point (pI), and allows for subsequent probing of immobilized complexes with specific antibodies. The method uses only minute amounts of sample, thus enabling the interrogation of primary specimens. Cultured cells were lysed in 20 mM HEPES pH 7.5, 50 mM KCl, 5 mM MgCl₂, 0.01% NP40, 20 mM Na₂MoO₄ buffer, containing protease and phosphatase inhibitors. Primary specimens were lysed in either Bicine-Chaps or RIPA buffers (ProteinSimple). Total protein assay was performed on an automated system, NanoPro 1000 Simple Western (ProteinSimple), for charge-based separation. Briefly, total cell lysates were diluted to a final protein concentration of 250 ng μ l⁻¹ using a master mix containing 1 \times Premix G2 pH 3–10 separation gradient (Protein simple) and 1 \times isoelectric point standard ladders (ProteinSimple). Samples diluted in this manner maintained their native charge state, and were loaded into capillaries (ProteinSimple) and separated based on their isoelectric points at a constant power of 21,000 μ Watts for 40 min. Immobilization was performed by UV-light embedded in the Simple Western system, followed by incubations with anti-HSP90 β (SMC-107A, StressMarq Biosciences), anti-HSP90 α (ab2928, Abcam), anti-HSP70 (SPA-810, Enzo), AKT (4691), P-AKT (9271) or BCL2 (2872) from Cell Signaling Technology and subsequently with HRP-conjugated anti-Mouse IgG (1030-05, SouthernBiotech) or with HRP-conjugated anti-Rabbit IgG (4010-05, SouthernBiotech). Protein signals were quantified by chemiluminescence using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific), and digital imaging and associated software (Compass) in the Simple Western system, resulting in a gel-like representation of the chromatogram. This representation is shown for each figure.

Western blotting. Protein was extracted from cultured cells in 20 mM Tris pH 7.4, 150 mM NaCl, 1% NP-40 buffer with protease and phosphatase inhibitors added (Complete tablets and PhosSTOP EASYpack, Roche). Ten to fifty μ g of total protein was subjected to SDS-PAGE, transferred onto nitrocellulose membrane, and incubated with indicated antibodies. HSP90 β (SMC-107) and HSP110 (SPC-195) antibodies were purchased from Stressmarq; HER2 (28-0004) from Zymed; HSP70 (SPA-810), HSC70 (SPA-815), HIP (SPA-766), HOP (SRA-1500), and HSP40 (SPA-400) from Enzo; HSP90 β (ab2927), HSP90 α (ab2928), p23 (ab2814), GAPDH (ab8245) and AHA1 (ab56721) from Abcam; cleaved PARP (G734A) from Promega; CDC37 (4793), CHIP (2080), EGFR (4267), S6K (2217), phospho-S6K (S235/236) (4858), P-AKT (S473) (9271), AKT (4691), P-ERK (T202/Y204) (4377),

ERK (4695), MCL1 (5453), Bcl-XL (2764), BCL2 (2872), c-MYC (5605) and HER3 (4754) from Cell Signaling Technology; and β -actin (A1978) from Sigma-Aldrich. The blots were washed with TBS/0.1% Tween 20 and incubated with appropriate HRP-conjugated secondary antibodies. Chemiluminescent signal was detected with Enhanced Chemiluminescence Detection System (GE Healthcare) following the manufacturer's instructions.

Native-cognate antibodies. We screened a panel of anti-chaperome antibodies for those that interacted with the target protein in its native form. We reasoned that these antibodies were more likely to capture stable multimeric forms of the chaperome members. These native-cognate antibodies were used in native-PAGE and IEF analyses of chaperome complexes. HSP90 β (SMC-107) and HSP110 (SPC-195) antibodies were purchased from Stressmarq; HSP70 (SPA-810), HSC70 (SPA-815), HOP (SRA-1500), and HSP40 (SPA-400) from Enzo; HSP90 β (ab2927), HSP90 α (ab2928), and AHA1 (ab56721) from Abcam; CDC37 (4793) from Cell Signaling Technology.

Native gel electrophoresis. Cells were lysed in 20 mM Tris pH 7.4, 20 mM KCl, 5 mM MgCl₂, 0.01% NP40, and 10% glycerol buffer by a freeze-thaw procedure. Primary samples were lysed in either Bicine-Chaps or RIPA buffers (ProteinSimple). Twenty-five to one hundred μ g of protein was loaded onto 4–10% native gradient gel and resolved at 4°C. The gels were immunoblotted as described above following either incubation in Tris-Glycine-SDS running buffer for 15 min before transfer in regular transfer buffer for 1 h, or directly transferred in 0.1% SDS-containing transfer buffer for 1 h.

siRNA knockdown. Cells were plated at 1×10^6 per 6 well-plate and transfected with an siRNA against human AHA1 (AHA1; 5'-TTCAAATGTGTCACGGATAA-3'), HSP90 α (HSP90AA1; no. 1 5'-ATGGCATGACAACACTCTTTAA-3'; no. 2 5'-AACCTGACCATTCATTATT-3'; no. 3 5'-TGCACTGTAAGACGTATGTAA-3'), HSP90 β (HSP90AB1; no. 1 5'-CAAGAATGATAAGGCAGTTAA-3'; no. 2 5'-TACGTTGCTCACTATTACGTA-3'; no. 3 5'-CAGAAGACAAGGAGAATTACA-3') HSP90 α/β (no. 1 5'-CAGAATGAAGGAGAACAGAA-3', no. 2 5'-CACAAACGATGATGAACAGTAT-3'), HSP110 (HSPH1; 5'-AGGCCGCTTTGTAGTTCAGAA-3') from Qiagen or HOP (STIP1) (Dharmacon; M-019802-01), or a negative control (scramble; 5'-CAGGGTATCGACGATTACAAA-3') with Lipofectamine RNAiMAX reagent (Invitrogen), incubated for 72 h and subjected to further analysis.

qRT-PCR. Total mRNA was isolated using TRIzol Reagent (Invitrogen) following the manufacturer's recommended protocol. Reverse transcription of mRNA into cDNA was performed using QuantiTect Reverse Transcription Kit (Qiagen). qRT-PCR was performed using PerfeCTa SYBR (Quanta Bioscience), 10 nM AHA1 (forward: 5'-GCGGCCGCTTCTAGTAGTTT-3' and reverse: 5'-CATCTCTCTCCGTCCAGTGC-3') and GAPDH (forward: 5'-CAAAGGCACAGTCAAGGCTGA-3' and reverse: 5'-TGGTGAAGACG CCACTAGATT-3') primers, or 1 \times QuantiTect Primers for HSP110 (HSPH1), HSP90 α (HSP90AA1), HSP90 β (HSP90AB1), HSP70 (HSPA1A), HOP (STIP1) (Qiagen) following recommended PCR cycling conditions. Melting curve analysis was performed to ensure product uniformity.

Protein depletion. To investigate which of the two HSP70 paralogues is involved in epichaperome formation we performed immunodepletions with HSP70 and HSC70 antibodies. Protein lysates were immunoprecipitated consecutively three times with either an HSP70 (Enzo, SPA-810), HSC70 (Enzo, SPA-815) or HOP (kindly provided by M. B. Cox, University of Texas at El Paso), or with the same species normal antibody as a negative control (Santa Cruz). The resulting supernatant was collected and run on a native or a denaturing gel.

Native gel electrophoresis and isoelectric focusing (IEF) under denaturing conditions. Tumour lysates were mixed with 10 M urea (dissolved in Felts buffer) to reach the indicated final concentrations of 2 M, 4 M and 6 M. After incubation for 10 min at room temperature or frozen overnight at –80°C, the lysates were loaded onto 4–10% native gradient gel and resolved at 4°C or applied to the IEF capillary. The HSP90 β bands were detected by using antibody purchased from Stressmarq (SMC-107).

MYC knockdown by lentiviral-delivered shRNA. A lentiviral vector expressing the MYC shRNA, as previously described³⁶, was requested from Addgene (Plasmid 29435, c-MYC shRNA sequence: GACGAGAACAGTTGAAACA). Viruses were prepared by co-transfecting the shRNA vector, the packaging plasmid psPAX2 and the envelop plasmid pMD2.G into HEK293 cells. OCI-LY1 cells were then infected with lentiviral supernatants in the presence of 4 μ g ml⁻¹ polybrene for 24 h. Following flow cytometry selection for positive cells, cells were expanded for further experiments. The MYC protein level was confirmed at 10 days post-infection by western blot using the anti-MYC antibody (Cell Signaling Technology, 5605).

Exogenous MYC expression. Viruses were prepared by co-transfection of the lentiviral vector expressing the MYC shRNA with pLM-mCerulean-2A-cMyc (Addgene, 23244) or pCDH-puro-cMYC (Addgene, 46970), the packaging plasmid psPAX2, and the envelope plasmid pMD2.G into HEK293 cells. ASPC1 cells were

then infected with lentiviral supernatants in the presence of $4\mu\text{g ml}^{-1}$ polybrene for 24 h and sorted for mCerulean positive cells or selected with puromycin treatment. Changes in cell size after infection were monitored by analysing the forward scatter (FSC) of intact cells via flow cytometry. MYC protein levels were analysed at 4 days post-infection by western blot.

MYC transcription factor binding activity assay. Whole cell extracts were prepared by homogenizing cells in RIPA buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1% NP40, 0.25% sodium deoxycholate, 10% glycerol, protease inhibitors). MYC activity was determined using the TransAM c-Myc Kit (Active Motif, 43396), following the manufacturer's instructions.

Cell viability assessment ATP assay. Cell viability was assessed using CellTiter-Glo luminescent Cell Viability Assay (Promega) after a 72 h PU-H71 treatment. The method determines the number of viable cells in culture based on quantification of the ATP present, which signals the presence of metabolically active cells, and was performed as previously reported³⁷. For the annexin V staining, cells were labelled with Annexin V-PE and 7AAD after PU-H71 treatment for 48 h, as previously reported³⁸. The necrotic cells were defined as annexin V⁺/7AAD⁺, and the early apoptotic cells were defined as annexin V⁺/7AAD⁻. For the LDH assay the release of lactate dehydrogenase (LDH) into the culture medium only occurs upon cell death. Following indicated treatment, the culture medium was collected and centrifuged to remove living cells and cell debris. The collected medium was incubated at room temperature for 30 min with the Cytotox-96 Non-radioactive Assay kit (Promega) LDH substrate.

In vivo studies. All animal studies were conducted in compliance with MSKCC's Institutional Animal Care and Use Committee (IACUC) guidelines. Female athymic nu/nu mice (NCRNU-M, 20–25 g, 6 weeks old) were obtained from Harlan Laboratories and were allowed to acclimatize at the MSKCC vivarium for 1 week before implanting tumours. Mice were provided with food and water *ad libitum*. Tumour xenografts were established on the forelimbs for PET imaging and on the flank for efficacy studies. Tumours were initiated by sub-cutaneous injection of 1×10^7 cells for MDA-MB-468 and 5×10^6 for ASPC1 in a 200 μl cell suspension of a 1:1 v/v mixture of PBS with reconstituted basement membrane (BD Matrigel, Collaborative Biomedical Products). Before administration, a solution of PU-H71 was formulated in citrate buffer. Sample size was chosen empirically based on published data³⁹. No statistical methods were used to predetermine sample size. Animals were randomly assigned to groups. Studies were not conducted blinded.

Small-animal PET imaging. Imaging was performed with a dedicated small-animal PET scanner (Focus 120 microPET; Concorde Microsystems, Knoxville, TN). Mice were maintained under 2% isoflurane (Baxter Healthcare, Deerfield, IL) anaesthesia in oxygen at 2 litres per min during the entire scanning period. To reduce the thyroid uptake of free iodide arising from metabolism of tracer, mice received 0.01% potassium iodide solution in their drinking water starting 48 h before tracer administration. For PET imaging, each mouse was administered 9.25 MBq (250 μCi) of ^{124}I -PU-H71 via the tail vein. List-mode data (10 to 30 min acquisitions) were obtained for each animal at various time points post-tracer administration. An energy window of 420–580 keV and a coincidence timing window of 6 ns were used. The resulting list-mode data were sorted into 2-dimensional histograms by Fourier rebinning; transverse images were reconstructed by filtered back projection (FBP). The image data were corrected for non-uniformity of scanner response, dead-time count losses, and physical decay to the time of injection. There was no correction applied for attenuation, scatter, or partial-volume averaging. The measured reconstructed spatial resolution of the Focus 120 is 1.6-mm FWHM at the centre of the field of view. Region of interest (ROI) analysis of the reconstructed images was performed using ASIPro software (Concorde Microsystems, Knoxville, TN), and the maximum pixel value was recorded for each tissue/organ ROI. A system calibration factor (that is, $\mu\text{Ci per ml per cps per voxel}$) that was derived from reconstructed images of a mouse-size water-filled cylinder containing ^{18}F was used to convert the ^{124}I voxel count rates to activity concentrations (after adjustment for the ^{124}I positron branching ratio). The resulting image data were then normalized to the administered activity to parameterize the micro-PET images in terms of per cent injected dose per gram (%ID per g) (corrected for decay of ^{124}I to the time of injection). Post-reconstruction smoothing was applied only for visual representation of images in the figures. Upon euthanasia, radioactivity (^{124}I) was measured in a gamma-counter (Perkin Elmer 1480 Wizard 3 Auto Gamma counter) using a 400–600 keV energy window. Count data were background- and decay-corrected to the time of injection, and the percent injected dose per gram (%ID per g) for each tumour sample was calculated using a calibration curve to convert counts to radioactivity, followed by normalization to the total activity injected.

Efficacy studies. Mice ($n = 5$) bearing MDA-MB-468 or ASPC1 tumours reaching a volume of 100–150 mm^3 were treated i.p. using PU-H71 (75 mg per kg) or vehicle, on a 3 times per week schedule, as indicated. Tumour volume (in mm^3)

was determined by measurement with Vernier calipers, and was calculated as the product of its length \times width² \times 0.5. Tumour volume was expressed on indicated days as the median tumour volume \pm s.d. indicated for groups of mice. Mice were euthanized after similar PU-H71 treatment periods, and at a time before tumours reached a size that resulted in discomfort or difficulty in physiological functions of mice in the individual treatment group, in accordance with our IUCAC protocol. **LC-MS/MS analyses.** Frozen tissue was dried and weighed before homogenization in acetonitrile/ H_2O (3:7). PU-H71 was extracted in methylene chloride, and the organic layer was separated and dried under vacuum. Samples were reconstituted in mobile phase. The concentrations of PU-H71 in tissue or plasma were determined by high-performance LC-MS/MS. PU-H71- d_6 was added as the internal standard⁴⁰. Compound analysis was performed on the 6410 LC-MS/MS system (Agilent Technologies) in multiple reaction monitoring mode using positive-ion electrospray ionization. For tissue samples, a Zorbax Eclipse XDB-C18 column (2.1 \times 50 mm, 3.5 μm) was used for the LC separation, and the analyte was eluted under an isocratic condition (80% H_2O + 0.1% HCOOH : 20% CH_3CN) for 3 min at a flow rate of 0.4 ml min^{-1} . For plasma samples, a Zorbax Eclipse XDB-C18 column (4.6 \times 50 mm, 5 μm) was used for the LC separation, and the analyte was eluted under a gradient condition (H_2O + 0.1% HCOOH : CH_3CN , 95:5 to 70:30) at a flow rate of 0.35 ml min^{-1} .

Chemical bait precipitation and proteomics. Protein extracts were prepared either in 20 mM HEPES pH 7.5, 50 mM KCl, 5 mM MgCl_2 , 1% NP40, and 20 mM Na_2MoO_4 for PU-H71 beads pull-down, or in 20 mM Tris pH 7.4, 150 mM NaCl, and 1% NP40 for YK beads pull-down. Samples were incubated with the PU-H71 beads (HSP90 bait) for 3–4 h or with the YK beads (HSP70 bait, for chemical precipitation) overnight, at 4 °C, then washed and subjected to SDS-PAGE with subsequent immunoblotting and western blot analysis. For HSP70 proteomic analyses, cells were incubated with a biotinylated YK-derivative, YK-biotin. Briefly, MDA-MB-468 cells were treated for 4 h with 100 μM biotin-YK5 or D-biotin as a negative control. Cells were collected and lysed in 20 mM Tris pH 7.4, 150 mM NaCl, and 1% NP40 buffer. Protein extracts were incubated with streptavidin agarose beads (Thermo Scientific) for 1 h at 4 °C, washed with 20 mM Tris pH 7.4, 150 mM NaCl, and 0.1% NP40 buffer and applied onto SDS-PAGE. The gels were stained with SimplyBlue Coomassie stain (Invitrogen Life Science Technologies). Proteomic analyses were performed using the published protocol^{7,18,22}. Control beads contained an inert molecule as previously described^{7,18,22}.

Protein identification by nano-liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analysis. Affinity-purified protein complexes from type 1 tumours ($n = 6$; NCI-H1975, MDA-MB-468, OCI-LY1, Daudi, IBL1, BC3), type 2 tumours ($n = 3$; ASPC1, OCI-LY4, Ramos) and from non-transformed cells ($n = 3$; MRC5, HMEC and neurons) were resolved using SDS-polyacrylamide gel electrophoresis, followed by staining with colloidal, SimplyBlue Coomassie stain (Invitrogen Life Science Technologies) and excision of the separated protein bands. Control beads that contained an inert molecule were subjected to the same steps as PU-H71 and YK beads and served as a control experiment. To ensure that we captured a majority of the HSP90 complexes in each cell type, we performed these studies under conditions of HSP90-bait saturation. The number of gel sections per lane averaged to be 14. *In situ* trypsin digestion of gel bound proteins, purification of the generated peptides and LC-MS/MS analysis were performed using our published protocols^{7,18,22}. After the acquisition of raw files, Proteowizard (version 3.0.3650)⁴¹ was used to create a Mascot Generic Format (mgf) file containing accurate mass for each peak and its corresponding ms^2 ions. Each mgf was then subjected to search a human segment of Uniprot protein database (20,273 sequences, European Bioinformatics Institute, Swiss Institute of Bioinformatics and Protein Information Resource) using Mascot (Matrix Science; version 2.5.0; <http://www.matrixscience.com>). Decoy proteins were added to the search to allow for the calculation of false discovery rates (FDR). The search parameters were as follows: (i) two missed cleavage tryptic sites were allowed; (ii) precursor ion mass tolerance = 10 p.p.m.; (iii) fragment ion mass tolerance = 0.8 Da; and (iv) variable protein modifications were allowed for methionine oxidation, deamidation of asparagine and glutamines, cysteine acrylamide derivatization and protein N-terminal acetylation. MudPit scoring was typically applied using significance threshold score $P < 0.01$. Decoy database search was always activated and, in general, for merged LS-MS/MS analysis of a gel lane with $P < 0.01$, false discovery rate averaged around 1%. The Mascot search result was finally imported into Scaffold (Proteome Software, Inc.; version 4.4.1) to further analyse tandem mass spectrometry (MS/MS) based protein and peptide identifications. X! Tandem (The GPM, <http://thegpm.org>; version CYCLONE (2010.12.01.1) was then performed and its results are merged with those from Mascot. The two search engine results were combined and displayed at 1% FDR. Protein and peptide probability was set at 95% with a minimum peptide requirement of 1. Protein identifications were expressed as Exclusive Spectrum Counts that identified each

protein listed. Primary data, such as raw mass spectrometry files, Mascot generic format files and proteomics data files created by Scaffold have been deposited onto the MassIVE site (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>; MassIVE Accession ID: MSV000079877). In each of the Scaffold files that validate and import Mascot searched files, peptide matches, scoring information (Mascot, as well as X! Tandem search scores) for peptide and protein identifications, MS/MS spectra, protein views with sequence coverage and more, can be easily accessed. To read the Scaffold files, free viewer software can be found at (<http://www.proteome-software.com/products/free-viewer/>). Peptide matches and scoring information that demonstrate the data processing are available in Supplementary Table 1f–q.

Bioinformatics analyses. The exclusive spectrum count values, an alternative for quantitative proteomic measurements⁴², were used for protein analyses. CHIP and PP5 were examined and used as internal quality controls among the samples. Statistics were performed using R (version 3.1.3) limma package^{43,44}. For entries with zero spectral counts, and to enable further analyses, we assigned an arbitrary small number of 0.1. The data were then transformed into logarithmic base 10 for analysis. Linear models were fit to the transformed data and moderated standard errors were calculated using empirical Bayesian methods. For Fig. 1f and Extended Data Fig. 5a, a moderated *t*-statistic was used to compare protein enrichment between type 1 cells and combined type 2 and non-transformed cells⁴⁵. For Extended Data Fig. 5b, the *t*-statistic was performed to compare protein enrichment among type 1 cells, type 2 cells and non-transformed cells (see Supplementary Table 1). Heat maps were created to display the selected proteins using the package “gplots” and “lattice”^{46,47}. See Supplementary Table 1 in which the table tab ‘a’ corresponds to Fig. 1f and contains core chaperome networks in type 1, type 2 and non-transformed cells; the table tab ‘b’ corresponds to Extended Data Fig. 5a and contains comprehensive chaperome networks in type 1, type 2 and non-transformed cells; the table tab ‘c’ corresponds to Extended Data Fig. 5b and Extended Data Fig. 8b and contains the HSP90 interactome as isolated by the HSP90 bait in type 1, type 2 and non-transformed cells; the table tab ‘d’ corresponds to Extended Data Fig. 8a and contains upstream transcriptional regulators that explain the protein signature of type 1 tumours and the table tab ‘e’ contains metastasis-related proteins characteristic of type 1 tumours.

The protein–protein interaction (PPI) network and upstream transcriptional regulators. To understand the physical and functional protein–interaction properties of the HSP90-interacting chaperome proteins enriched in type 1 tumours, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database⁴⁸. Proteins displayed in the heat map were uploaded in STRING database to generate the PPI networks. STRING builds functional protein–association networks based on compiled available experimental evidence. The thickness of the edges represents the confidence score of a functional association. The score was calculated based on four criteria: co-expression, experimental and biochemical validation, association in curated databases, and co-mentioning in PubMed abstracts⁴⁸. Proteins with no adjacent interactions were not shown. The colour scale in nodes indicates the average enrichment of the protein (measured as exclusive spectral counts) in type 1, type 2, and non-transformed cells, respectively. The network layout for type 1 tumours was generated using edge-weighted spring–electric layout in Cytoscape with slight adjustments of marginal nodes for better visualization⁴⁹. The layout for type 2 and non-transformed cells retains that of type 1 for better comparison. Proteins with average relative abundance values less than 1 were deleted from analyses. The biological processes in which they participate and the functionality of proteins enriched in type 1 tumours were assigned based on gene ontology terms and based on their designated interactome from UniProtKB, STRING, and/or I2D databases^{48,50–53}. The Upstream Regulator analytic, as implemented in Ingenuity Pathways Analysis (IPA, QIAGEN Redwood City, <http://www.qiagen.com/ingenuity>), was used to identify the cascade of upstream transcriptional regulators that can explain the observed protein expression changes in type 1 tumours. The analysis is based on prior knowledge of expected effects between transcriptional regulators and their target genes stored in the Ingenuity Knowledge Base. The analysis examines how many known targets of each transcription regulator are present in the data set, and calculates an overlap *P* value for upstream regulators based on significant overlap between dataset genes and known targets regulated by a transcription regulator. For Extended Data Fig. 8b, proteins were selected based on 3 pre-curated lists (MYC target genes based on the analysis report from INGENUITY, MYC signature genes based on the reported list provided in ref. 54 and MYC expression/function activators were manually curated from UniProt and GeneCards databases).

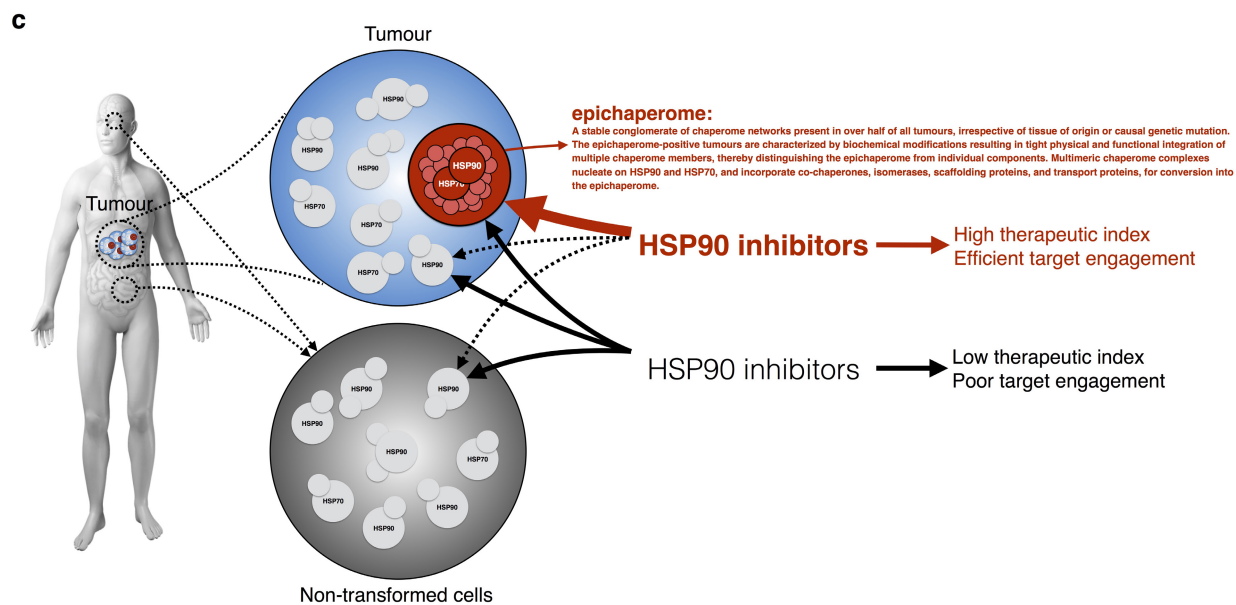
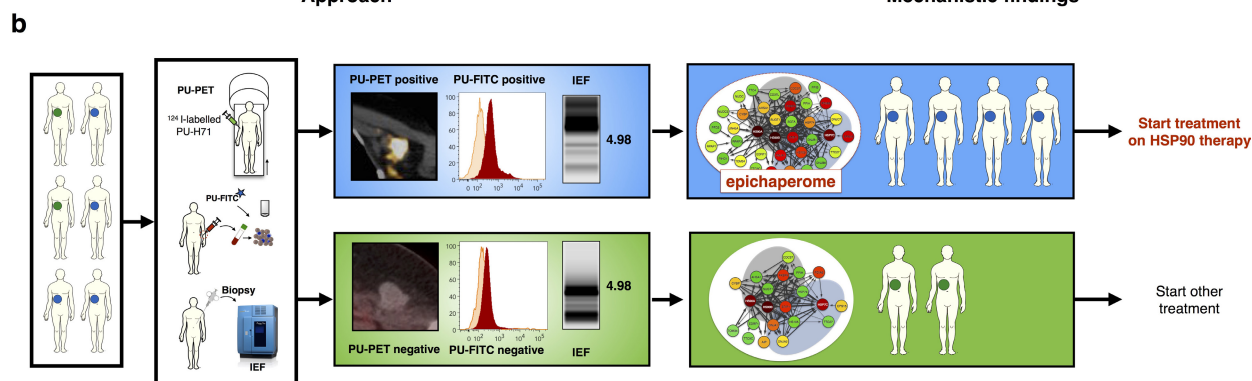
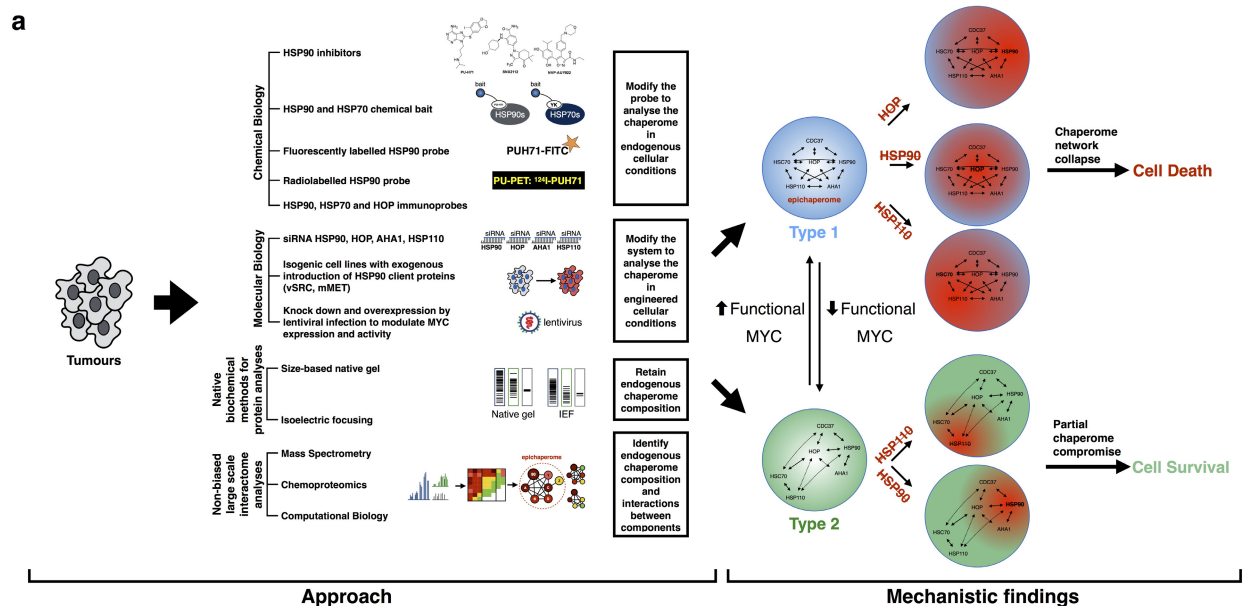
Sequencing data. Cell lines with information available in the cBioPortal for cancer genomics (<http://www.cbioportal.org>) were evaluated for mutations in pathways implicated in cancer: *P53*, *RAS*, *RAF*, *PTEN*, *PIK3CA*, *AKT*, *EGFR*, *HER2*, *CDK2NA/B*, *RB*, *MYC*, *STAT1*, *STAT3*, *JAK2*, *MET*, *PDGFR*, *KDM6A*,

KIT. Mutations in major chaperome members (*HSP90AA1*, *HSP90AB1*, *HSPH1*, *HSPA8*, *STIP1*, *AHSA1*) were also evaluated.

Statistical analysis. Data were visualized and statistical analyses performed using GraphPad Prism (version 6; GraphPad Software) or R statistical package. In each group of data, estimate variation was taken into account and is indicated in each figure as s.d. or s.e.m. If a single panel is presented, data are representative of 2 or 3 biological or technical replicates, as indicated. *P* values for unpaired comparisons between two groups with comparable variance were calculated by two-tailed Student's *t*-test. Pearson's tests were used to identify correlations among variables. Significance for all statistical tests was shown in figures for not significant (NS), **P* < 0.05, ***P* < 0.01, ****P* < 0.001 and *****P* < 0.0001. No samples or animals were excluded from analysis, and sample size estimates were not used. Animals were randomly assigned to groups. Studies were not conducted blinded, with the exception of all patient specimen histological analyses.

- Taldone, T. *et al.* Design, synthesis, and evaluation of small molecule Hsp90 probes. *Bioorg. Med. Chem.* **19**, 2603–2614 (2011).
- Rodina, A. *et al.* Affinity purification probes of potential use to investigate the endogenous Hsp70 interactome in cancer. *ACS Chem. Biol.* **9**, 1698–1705 (2014).
- Taldone, T. *et al.* Synthesis of purine-scaffold fluorescent probes for heat shock protein 90 with use in flow cytometry and fluorescence microscopy. *Bioorg. Med. Chem. Lett.* **21**, 5347–5352 (2011).
- Chiosis, G. *et al.* Uses of labeled hsp90 inhibitors. US patent 20140242602 A1 (2014).
- Taldone, T. *et al.* Radiosynthesis of the iodine-124 labeled Hsp90 inhibitor PU-H71. *J. Labelled Comp. Radiopharm.* **59**, 129–132 (2016).
- Rodina, A. *et al.* Identification of an allosteric pocket on human Hsp70 reveals a mode of inhibition of this therapeutically important protein. *Chem. Biol.* **20**, 1469–1480 (2013).
- Taldone, T. *et al.* Heat shock protein 70 inhibitors. 2. 2,5'-thiodipyrimidines, 5-(phenylthio)pyrimidines, 2-(pyridin-3-ylthio)pyrimidines, and 3-(phenylthio)pyridines as reversible binders to an allosteric site on heat shock protein 70. *J. Med. Chem.* **57**, 1208–1224 (2014).
- Kang, Y. *et al.* Heat shock protein 70 inhibitors. 1. 2,5'-thiodipyrimidine and 5-(phenylthio)pyrimidine acrylamides as irreversible binders to an allosteric site on heat shock protein 70. *J. Med. Chem.* **57**, 1188–1207 (2014).
- Miyajima, N. *et al.* The HSP90 inhibitor ganetespib synergizes with the MET kinase inhibitor crizotinib in both crizotinib-sensitive and -resistant MET-driven tumor models. *Cancer Res.* **73**, 7022–7033 (2013).
- Whitesell, L., Mimnaugh, E. G., De Costa, B., Myers, C. E. & Neckers, L. M. Inhibition of heat shock protein HSP90-pp60^{v-src} heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. *Proc. Natl Acad. Sci. USA* **91**, 8324–8328 (1994).
- Corben, A. D. *et al.* *Ex vivo* treatment response of primary tumors and/or associated metastases for preclinical and clinical development of therapeutics. *J. Vis. Exp.* **92**, e52157 (2014).
- Hassane, D. C. *et al.* Chemical genomic screening reveals synergism between parthenolide and inhibitors of the PI-3 kinase and mTOR pathways. *Blood* **116**, 5983–5990 (2010).
- Miller, J. D. *et al.* Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* **13**, 691–705 (2013).
- Moullick, K. *et al.* Synthesis of a red-shifted fluorescence polarization probe for Hsp90. *Bioorg. Med. Chem. Lett.* **16**, 4515–4518 (2006).
- Fan, A. C. *et al.* Nanofluidic proteomic assay for serial analysis of oncoprotein activation in clinical specimens. *Nat. Med.* **15**, 566–571 (2009).
- Lin, C. H., Jackson, A. L., Guo, J., Linsley, P. S. & Eisenman, R. N. Myc-regulated microRNAs attenuate embryonic stem cell differentiation. *EMBO J.* **28**, 3157–3170 (2009).
- Rodina, A. *et al.* Selective compounds define Hsp90 as a major inhibitor of apoptosis in small-cell lung cancer. *Nat. Chem. Biol.* **3**, 498–507 (2007).
- Zong, H. *et al.* A hyperactive signalosome in acute myeloid leukemia drives addition to a tumor-specific Hsp90 species. *Cell Reports* **13**, 2159–2173 (2015).
- Caldas-Lopes, E. *et al.* Hsp90 inhibitor PU-H71, a multimodal inhibitor of malignancy, induces complete responses in triple-negative breast cancer models. *Proc. Natl Acad. Sci. USA* **106**, 8368–8373 (2009).
- Taldone, T., Zatorska, D., Kang, Y. & Chiosis, G. A facile and efficient synthesis of d6-labeled PU-H71, a purine-scaffold Hsp90 inhibitor. *J. Labelled Comp. Radiopharm.* **53**, 47–49 (2010).
- Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536 (2008).
- Skarra, D. V. *et al.* Label-free quantitative proteomics and SAINT analysis enable interactome mapping for the human Ser/Thr protein phosphatase 5. *Proteomics* **11**, 1508–1516 (2011).
- R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2013).
- Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

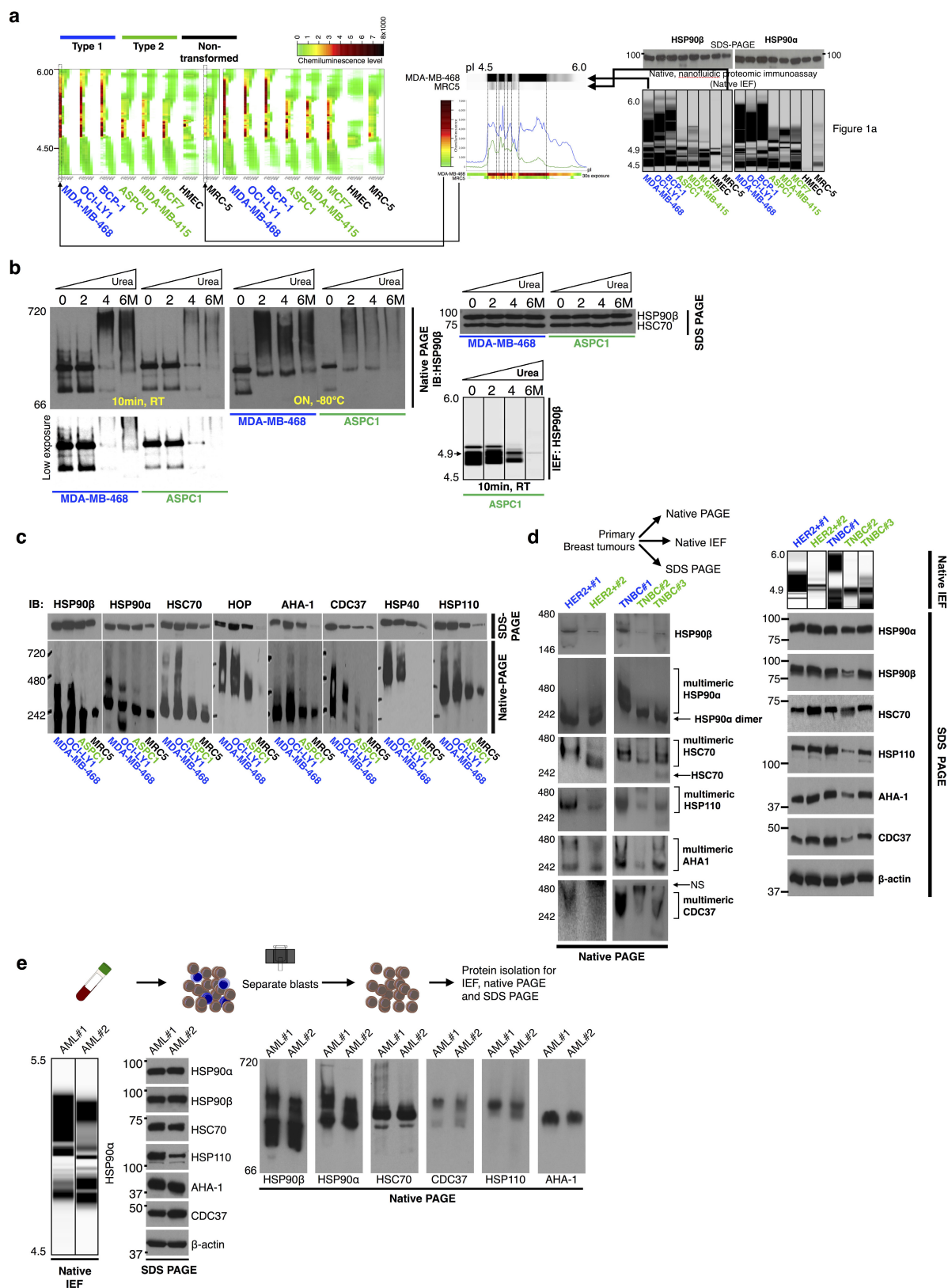
45. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
46. Sarkar, D. *Lattice: Multivariate Data Visualization with R*. (Springer, 2008).
47. Scales, M., Jäger, R., Migliorini, G., Houlston, R. S. & Henrion, M. Y. visPig—a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. *PLoS One* **9**, e107497 (2014).
48. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
49. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
50. Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **40**, D857–D861 (2012).
51. Brown, K. R. & Jurisica, I. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.* **8**, R95 (2007).
52. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
53. Brown, K. R. & Jurisica, I. Online predicted human interaction database. *Bioinformatics* **21**, 2076–2082 (2005).
54. Chandriani, S. *et al.* A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS One* **4**, e6693 (2009).



Extended Data Figure 1 | See next page for caption.

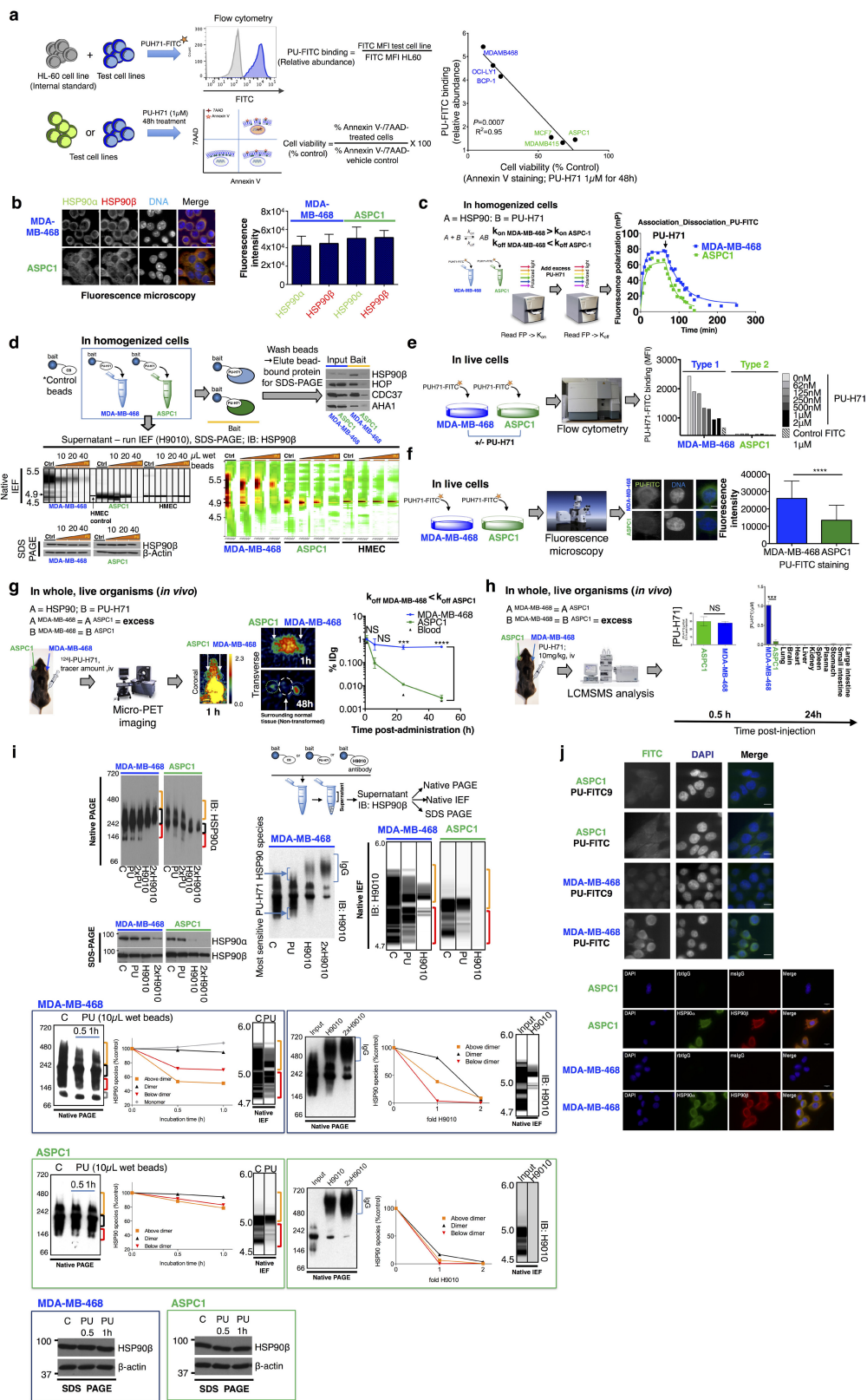
Extended Data Figure 1 | Summary of the experimental design and findings. **a**, Schematic of the experimental approach to address four key questions concerning the chaperome in cancer: (1) what are the molecular characteristics and composition of the chaperome in cancer; (2) what molecular factors drive chaperome networks to crosstalk in tumours; (3) what distinguishes the chaperomes of tumours that are sensitive to pharmacologic inhibition from those that are not; and (4) what are the characteristics of tumours that may benefit from chaperome therapy? To retain the endogenous proteome/chaperome make-up and function, we applied a variety of chemical biology tools and biochemical methods that retain native protein conformations and complexes. This approach minimally interferes with the system it interrogates, thus providing answers closer to the reality of disease. When applicable, data were validated by alternative and complementary methods, as indicated. This approach led to the discovery of a novel mechanism of tumour regulation. Specifically, we have identified and characterized the epichaperome, a modified chaperome network. Our data demonstrate that heterogeneous and stable, multimeric chaperome complexes nucleating on HSP90 and HSP70, and incorporating co-chaperones, isomerases, scaffolding proteins, and transport proteins, bring about the effective physical and functional integration of the chaperome machinery into the epichaperome. Chaperome rewiring into the epichaperome is fuelled by powerful transcription activators such as MYC. Only under conditions in which the chaperome becomes tightly integrated both functionally and

physically to form the epichaperome are tumours addicted to individual chaperome members. The epichaperome is the survival mechanism for type 1 tumours; when the epichaperome is dismantled by ablation of a component chaperome, the chaperome network collapses leading to cell death. In contrast, in type 2 tumours in which the integration of the chaperome is only partial and most chaperome members function as insular communities, depletion of chaperome members only 'locally' compromises the chaperome, maintaining overall cellular survival. **b, c**, Therapeutic and diagnostic implications of the findings. We propose the epichaperome as a biomarker to stratify patients for chaperome therapy, such as HSP90 inhibitors. This work also provides several ways to measure the epichaperome in clinic, that is, a non-invasive imaging assay (PU-PET) for solid tumours, a flow cytometry assay based on PU-FITC for liquid tumours and a native protein separation and analysis for minute biopsy specimens (isoelectric focusing; NanoPro technique) (**b**). We also propose that HSP90 is a cancer target when integrated into the epichaperome. Thus, HSP90 inhibitors that are specific for HSP90 when part of the epichaperome would be preferred for clinical use. Non-discriminate pharmacological agents that target chaperome members regardless of whether they are in the epichaperome or are part of dynamic complexes, such as in normal cells, could lead to toxicities and a low therapeutic index. For example, GI and ocular toxicities have been associated with some HSP90 inhibitors and not others due to chronic HSP90 inhibition in these normal tissues.



Extended Data Figure 2 | Biochemical profile of chaperome members in cancer cell lines and primary specimens. **a**, The biochemical profile of HSP90 in cell lines was analysed by native capillary isoelectric focusing. The ‘heat map’ representation shows snapshots of HSP90 complexes as detected under different antibody blotting exposure times. See also Fig. 1a. **b**, The biochemical profile of HSP90 in samples denatured with urea. RT, room temperature; ON, overnight. Data were repeated independently twice with representative data shown. **c–e**, The biochemical

profile of indicated chaperome members in cell lines (**c**, $n = 4$) and primary specimens (breast cancer, $n = 5$ (**d**), and acute myeloid leukaemia (AML), $n = 2$ (**e**)) was analysed by native capillary isoelectric focusing (IEF), native-PAGE and SDS–PAGE. The schematic for the isolation and separation of AML blasts for biochemical analyses is shown in **e**, top. TNBC, triple negative breast cancer; HER2⁺, HER2-overexpressing breast cancer. For uncropped gel data, see Supplementary Fig. 1.

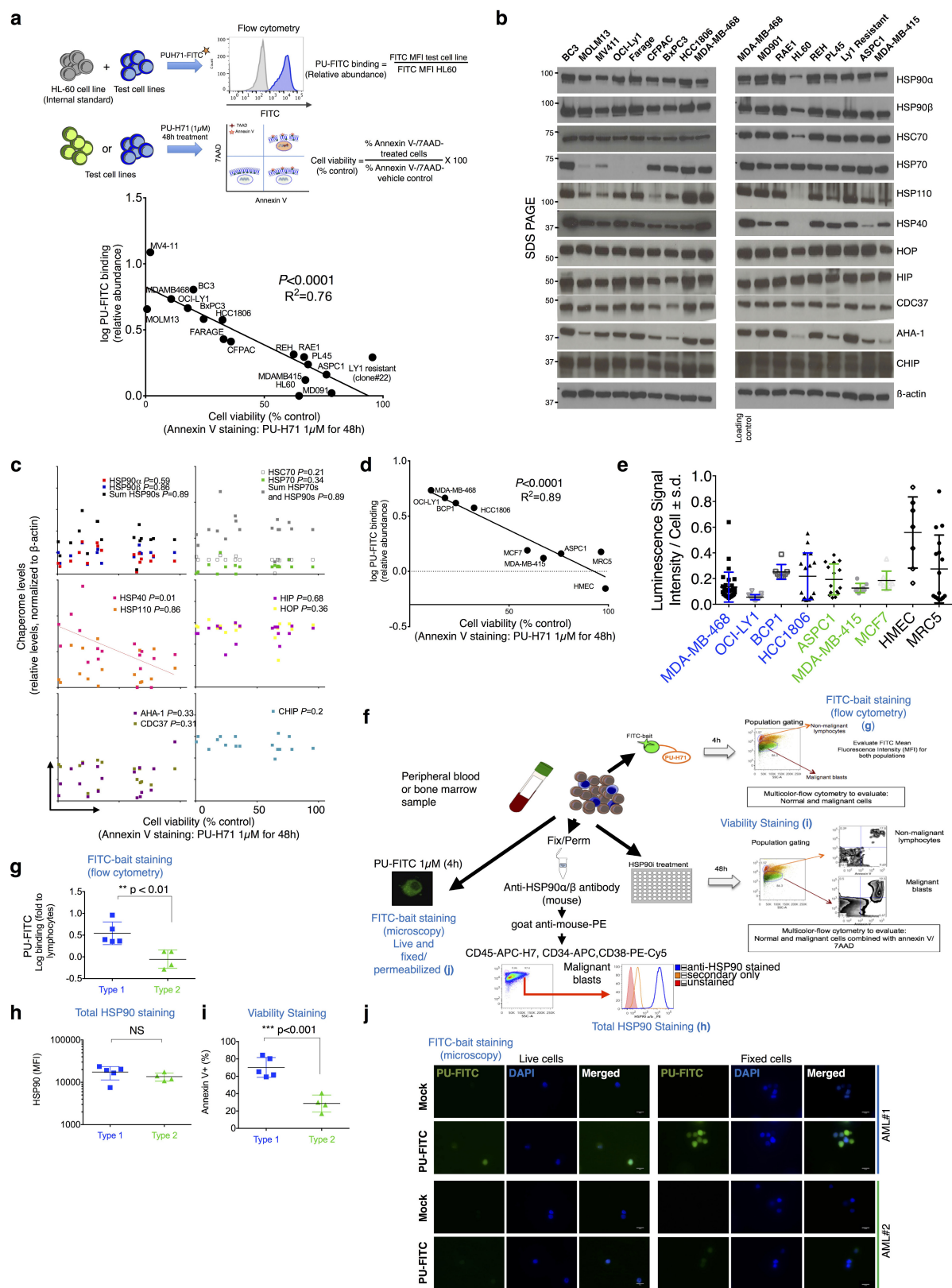


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | PU-H71 and its labelled versions are reliable tools to perturb, identify and measure the expression of the high-molecular-weight, multimeric HSP90 complexes in tumours.

a, Correlative analysis between binding of a fluorescently (FITC) labelled PU-H71 (PU-FITC) to the panel of cancer cells shown in Fig. 1a ($n = 6$) and their apoptotic sensitivity to HSP90 inhibition (Pearson's r , two-tailed). Each data point represents a cell line. **b**, MDA-MB-468 (type 1) and ASPC1 (type 2) contain similar levels of HSP90 but only MDA-MB-468 expresses the high-molecular-weight chaperone species (see also Fig. 1a). HSP90 α and HSP90 β levels were quantified by fluorescence microscopy ($n = 50$; mean \pm s.d.; unpaired t -test, NS). Scale bar, 10 μ m. **c**, Association and dissociation of PU-FITC (a FITC labelled PU-H71) from HSP90 was probed in cell homogenates by fluorescence polarization. Average from technical duplicates is shown on the graph. The experiment was carried out twice with similar results. **d**, Association and dissociation of PU-H71-bait (a solid-support immobilised PU-H71) from HSP90 was probed in cell homogenates by chemical precipitation followed by analyses of HSP90 in the supernatant and of HSP90 isolated on the solid support. A solid-support containing immobilized PU-H71 and an HSP90-inert molecule (control bait) were incubated with cell homogenates for 2 h. The bait-captured cargo was isolated and analysed by western blot (bait). The HSP90 species in the supernatant were separated and analysed as indicated. For isoelectric focusing, both the gel (for experimental duplicates) and the heat map representations of different exposure times are shown for each experimental condition. HMEC cells are shown for reference. Data were repeated independently twice with representative data shown. **e**, **f**, Association and dissociation of PU-H71 from type 1 and 2 tumours, measured in cells. Binding of PU-FITC to live cells was analysed by flow cytometry and fluorescence microscopy, as indicated. PU-FITC (1 μ M for flow and 5 μ M for microscopy) was added to cells and incubated for 4 h before fluorescence signal detection. To show specificity of binding, the signal was competed off in a dose-dependent manner with unlabelled PU-H71. Control FITC, a triethylene glycol labelled FITC. **e**, Mean from two technical replicates; **f**, mean \pm s.d., $n = 50$ individual cells, unpaired t -test, **** $P < 0.0001$. The fluorescence intensity of PU-FITC staining was quantified by ImageJ. Scale bar, 10 μ m. **g**, Association and dissociation of PU-H71 from type 1 and 2 tumours, measured *in vivo*. Biodistribution of 124 I-PU-H71 (a 124 I radiolabelled version of

PU-H71) was monitored live in tumour-bearing mice. Each mouse bears one xenografted MDA-MB-468 and one ASPC1 tumour, of similar volume, as indicated. Following intravenous (iv) injection of a tracer amount of the 124 I-PU-H71 agent, mice were monitored by micro-positron emission tomography (PET) imaging. Representative images taken at the indicated times post-injection are shown. Note that immediately after injection (1 h timepoint image), the agent is widely and uniformly distributed throughout the body and in each tumour. The off rate from type 1 tumours is slower compared to type 2 and non-transformed tissues (that is, distinct k_{off} from type 1 tumours versus type 2 tumours). The image is representative of five individual mice. In an independent experiment, radioactivity was measured in a gamma-counter upon mouse euthanasia and data were graphed to monitor the time-dependent distribution of PU-H71. Graph; radioactivity, measured as %IDg; injected dose per gram, versus time upon euthanasia (mean \pm s.d., $n = 8$, ASPC1; $n = 34$, MDA-MB-468, pooled experiments of mice bearing individual tumours). Means were compared by unpaired t -tests between MDA-MB-468 and ASPC1 at each time point (NS, not significant; **** $P < 0.001$; **** $P < 0.0001$). **h**, Same as in **g** for a therapeutic dose of injected PU-H71, as indicated. Levels of intact PU-H71 in the indicated tumours, tissues and plasma were determined by liquid chromatography tandem mass spectrometry (LC-MS/MS) in mice ($n = 5$) euthanized at the indicated times post-PU-H71 injection. Graph; mean \pm s.d., unpaired t -tests between MDA-MB-468 and ASPC1 (NS, not significant; **** $P < 0.001$). **i**, Dose- and time-dependent binding of PU-H71 and H9010 (an anti-HSP90 antibody) to HSP90 species expressed in type 1 and type 2 tumour cells. C, control beads containing an HSP90-inert molecule. PU, 10 μ l PU-H71 wet beads; 2 \times PU, 20 μ l PU-H71 wet beads; H9010; 2 μ l antibody immobilized on agarose beads. Because the IgG interferes with the HSP90 signal (see the high molecular smear in the native gels), native lysates were used for a control (input). Graph shows quantification of time-dependent changes in HSP90 species. **j**, Representative fluorescence microscopy images of live cells stained with PU-FITC (top) as compared to antibodies specific for HSP90 (bottom). rbtIgG, rabbit IgG control, mslgG, mouse IgG control. Scale bar, 10 μ m. Micrograph is representative of four captured images. For uncropped gel data, see Supplementary Fig. 1.

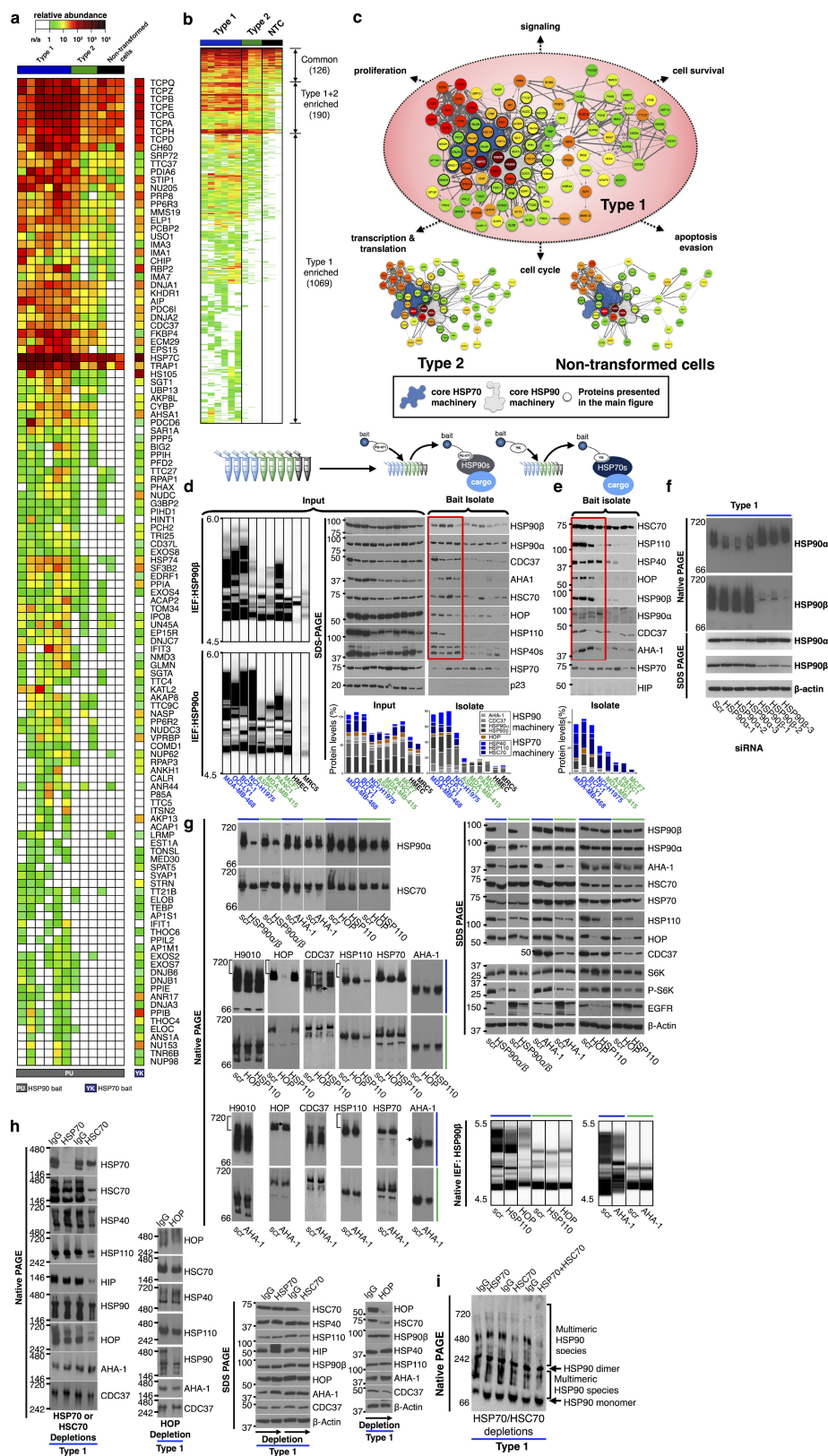


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Binding affinity of PU-H71 for cellular HSP90 is independent of the expression of HSP90 and other chaperome members, and is not affected by intracellular ATP concentration variations.

a, Correlative analysis for PU-H71-sensitive HSP90 species abundance, as measured by PU-FITC capture, and cell viability upon a 48 h treatment with PU-H71 (1 μ M), as measured by annexin V staining (Pearson's r , two-tailed). Each data point represents a cell line ($n = 17$); data points are the mean from two biological replicates ran in duplicate or triplicate. To account for intercellular background signal variability, HL60 cells were spiked in and used as internal control for each cell line; thus binding is presented as a ratio of the signal obtained in the analysed cell over that in HL60 cells. y axis, log values of the binding ratios. **b**, Cell lines analysed in **a** were lysed and total levels of the indicated chaperome members were determined by western blot. β -actin; protein loading control. **c**, A correlative analysis was performed between total chaperome levels, as obtained in **b**, and cell viability values, as determined in **a**; no significant and/or robust relationship was observed (Pearson's r , two-tailed). **d**, **e**, In a panel of type 1, type 2 and non-transformed cells ($n = 9$), binding to PU-H71 and cell killing by PU-H71 (**d**) was compared to intracellular ATP levels (**e**). **d**, Correlation, Pearson's r , two-tailed;

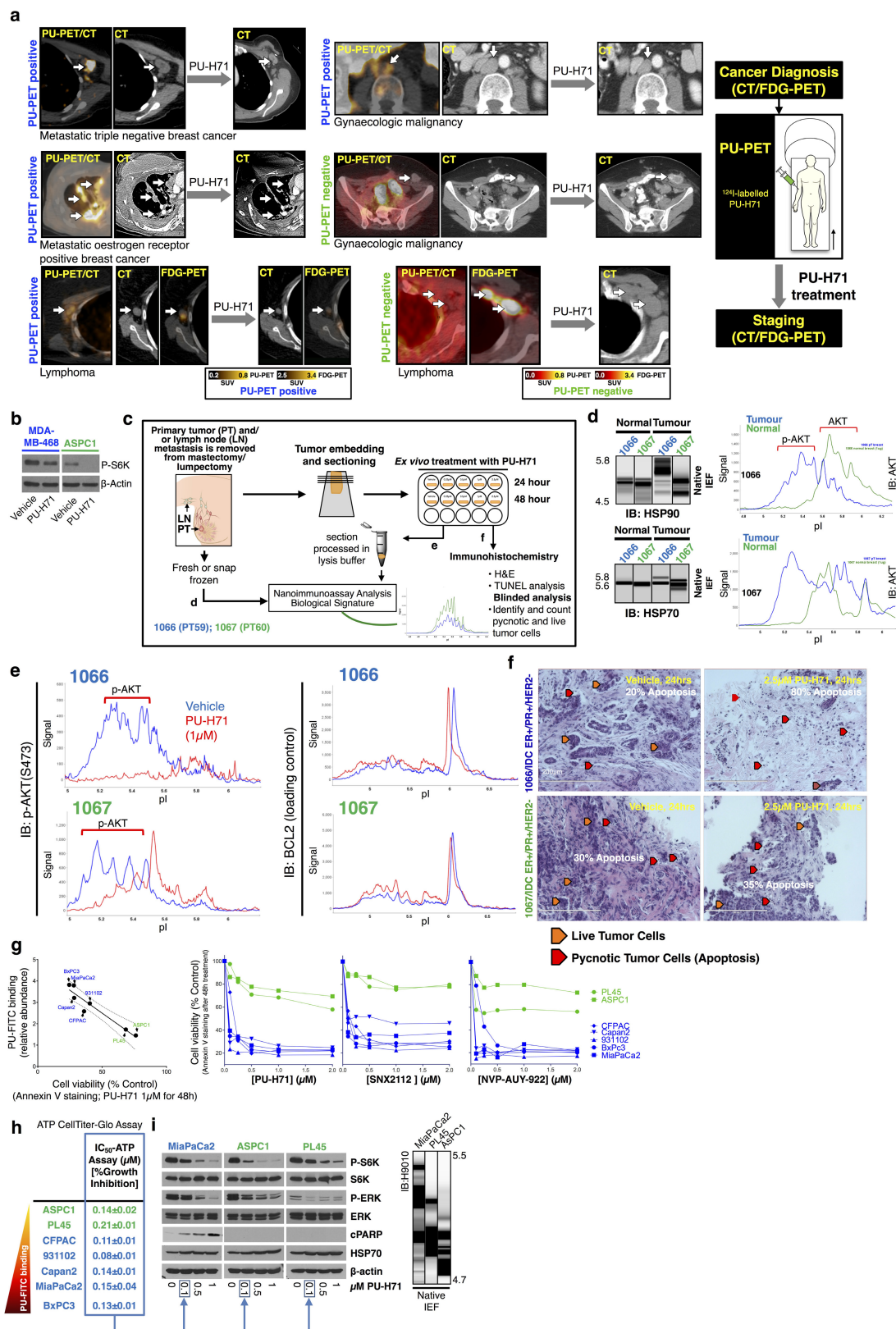
e, mean \pm s.d., each symbol represents an experimental replicate (MDA-MB-468, $n = 23$; OCI-LY1, $n = 15$; BCP-1, $n = 8$; HCC-1806, $n = 16$; ASPC1, $n = 15$; MDA-MB-415, $n = 8$; MCF7, $n = 8$; HMEC, $n = 7$; MRC5, $n = 16$). **f**, Schematic showing the experimental design for the isolation and analysis of primary AML samples. **g**, AML samples were stained with PU-FITC, and blasts (malignant) and lymphocytes (normal) were separated and analysed by flow cytometry. The signal in blasts over lymphocytes (used as internal standard) was graphed to classify clones as type 1 (>2.1 PU-FITC binding ratio of signal in blast versus lymphocytes) and type 2 (<2.1 PU-FITC binding ratio) (mean \pm s.d., $n = 9$, unpaired t -test, $**P < 0.01$). **h**, Total HSP90 levels were measured by staining with an HSP90 antibody after cell fixation and permeabilization (mean \pm s.d., $n = 9$, unpaired t -test, NS, not significant). **i**, Viability of blasts following exposure to PU-H71 (1 μ M) for 48 h was measured by annexin V/7AAD staining (mean \pm s.d., $n = 9$, unpaired t -test, $***P < 0.001$). **j**, PU-FITC staining of live and fixed/permeabilized unfractionated AMLs was visualized by fluorescence microscopy. Scale bar, 100 μ m. Micrograph is representative of two captured images. The biochemical profile of AML no. 1 and AML no. 2 is presented in Extended Data Fig. 2e. For uncropped gel data, see Supplementary Fig. 1.



Extended Data Figure 5 | See next page for caption.

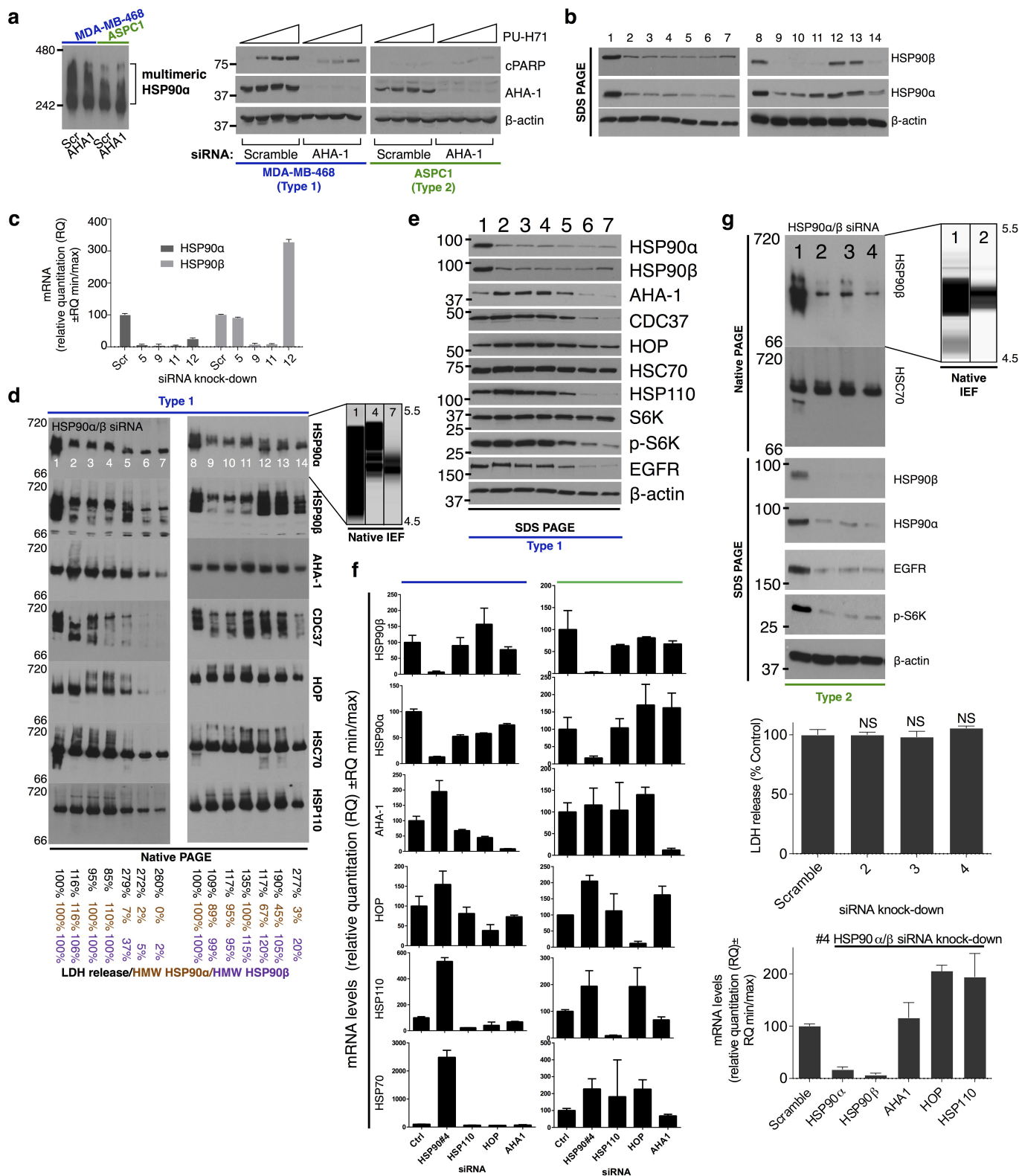
Extended Data Figure 5 | Chaperome networks in type 1, type 2 and non-transformed cells. **a, b,** Heat maps illustrating all chaperome members (**a**) and the interactome of HSP90 (**b**) isolated by the HSP90 bait and identified upon mass spectrometry and bioinformatics analyses enriched ($P < 0.1$) in type 1 tumours over type 2 tumours and non-transformed cells. Protein sorting was based on hierarchical clustering. Last lane of the heat map in **a** shows the enrichment of these proteins on the HSP70 bait. **c,** Network illustrating the connectivity of proteins isolated by the HSP90 bait and identified upon mass spectrometry and bioinformatics analyses; chaperome members and proteins with scaffolding, adaptor, protein interface modulator roles and significantly enriched ($P < 0.1$) in type 1 tumours over type 2 tumours and non-transformed cells are shown. The thickness of the edges (connection lines) represents the robustness of the functional interaction. The colour of nodes represents protein abundance. For comparison, type 2 and

non-transformed cells are also shown. Core interactions are shown in Fig. 1f. **d, e,** The cargo or interacting proteins of HSP90 (**d**) and HSP70 (**e**) isolated by the PU-H71 and YK-chemical baits from the indicated cell homogenates. Protein levels in individual cell homogenates (input) were analysed by IEF and SDS-PAGE, as indicated. Proteins precipitated on the chemical bait were analysed by SDS-PAGE. Protein levels from each experimental condition were quantified and graphed (bottom). Data were repeated independently twice with representative data shown. **f–i,** Changes in multimeric chaperone complexes in cells challenged with multiple siRNAs against HSP90 α or HSP90 β (**f**), HSP90 α/β , AHA1, HOP or HSP110 (**g**) and in cell homogenates challenged with antibodies specific for the indicated HSP70 paralogues and for HOP (**h, i**), as indicated. Levels of proteins in the homogenate were probed by SDS-PAGE or native-PAGE, as indicated. All data were repeated independently twice with representative data shown. For uncropped gels, see Supplementary Fig. 1.



Extended Data Figure 6 | HSP90 is functional and susceptible to exogenous inhibitors in type 2 as well as in type 1 cells, but only inhibition of HSP90 in type 1 cells is toxic to the cell. **a**, The response of type 1 and 2 tumours classified by PU-PET avidity, to PU-H71 treatment, is shown. Patients were treated as part of the NCT01393509 clinical study. Each picture is a scan of data taken of an individual patient. PU-PET images were taken at 24 h after ^{124}I -PU-H71 tracer administration. Scale bars (bottom of panel); PET window display intensity scales for FDG and PU-PET fusion PET-CT images. Numbers in the scale bars indicate upper and lower SUV thresholds that define pixel intensity on PET images. **b**, Changes in HSP90 machinery function upon pharmacologic inhibition (PU-H71, 1 μM for 24 h). Inhibition of PI3K/AKT activity was monitored; see p-S6K surrogate for AKT activity in cell lines and p-AKT in primary specimens (below). Data in cell lines were repeated independently twice with representative data shown. For HSP90 α/β knockdown data, see Extended Data Fig. 5g. **c–f**, Treatment schematic and representative examples of primary breast cancer specimens ($n = 2$) treated *ex vivo* with PU-H71. **c**, Workflow for the analysis of the primary specimens. **d**, Molecular signature of tumour and adjacent normal tissue of the surgical specimen as analysed by native, nanofluidic proteomic assay (NanoPro; native IEF), for HSP90 and HSP70 (gel representation), and AKT (chromatogram representation). **e**, Molecular response of tumour sections treated for 24 h *ex vivo* with PU-H71 (1 μM). AKT (an HSP90 client) activity was probed with the indicated antibody. BCL2 was chosen as a loading standard; this protein is insensitive to HSP90 inhibition in the analysed primary breast specimens (native IEF, chromatogram representation). **f**, Apoptotic response of the indicated tumour specimens to *ex vivo* treatment with PU-H71 or vehicle. Apoptosis and necrosis of the

tumour cells (as percentage) is assessed by reviewing all the haematoxylin and eosin (H&E) slides of the case (controls and treated ones) *in toto*, blindly, allowing for better estimation of the overall treatment effect to the tumour. Image representative of the entire specimen section. **g–j**, Response profile of a panel of pancreatic cancer cells to HSP90 inhibition. **g**, Changes in cell viability following HSP90 pharmacologic inhibition by three chemically distinct agents, as indicated. Mean from two to three technical replicates is shown. Subclassification of the analysed cell lines by PU-FITC binding is shown on the left. **h**, The effect of PU-H71 on cell growth was measured with an assay that analyses intracellular ATP levels. Cells were treated for 72 h with PU-H71 and the half maximal inhibitory growth concentration (IC_{50}) was determined. Mean \pm s.d.; $n = 6$. **i**, Representative examples of type 1 and type 2 cells treated for 24 h with the indicated concentrations of PU-H71. Inhibition of HSP90 is demonstrated by a decrease in HSP90 client function (p-S6K and p-ERK) and by HSP70 induction, and evidenced in both type 1 and 2 tumour cells. Induction of apoptosis, as demonstrated by the appearance of cleaved PARP (cPARP), is however, specific to type 1 tumour cells. β -actin, protein loading control. The HSP90 biochemical signature of the select cells is shown on the right. The blue arrows indicate the close relationship between the growth inhibitory IC_{50} values and HSP90 function inhibition, suggesting that HSP90 inactivation is sufficient to inhibit growth (that is, have a static effect) in both type 1 and 2 tumours. In contrast, substantial induction of apoptosis is specific to type 1 tumours. Thus, HSP90 is functional in type 2 and is engaged by the HSP90 inhibitors—the resistance phenotype of type 2 tumours cannot be explained by an inability of the HSP90 inhibitor to engage HSP90. Data are representative of two independent experiments. For uncropped gel data, see Supplementary Fig. 1.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | The epichaperome facilitates cancer

cell survival. The expression of the epichaperome was altered by epichaperome components knockdown (**a**) or by titrating into cells siRNAs that targeted both the HSP90 α and the HSP90 β paralogues (**b–g**) to test whether the epichaperome facilitates survival in type 1 tumours.

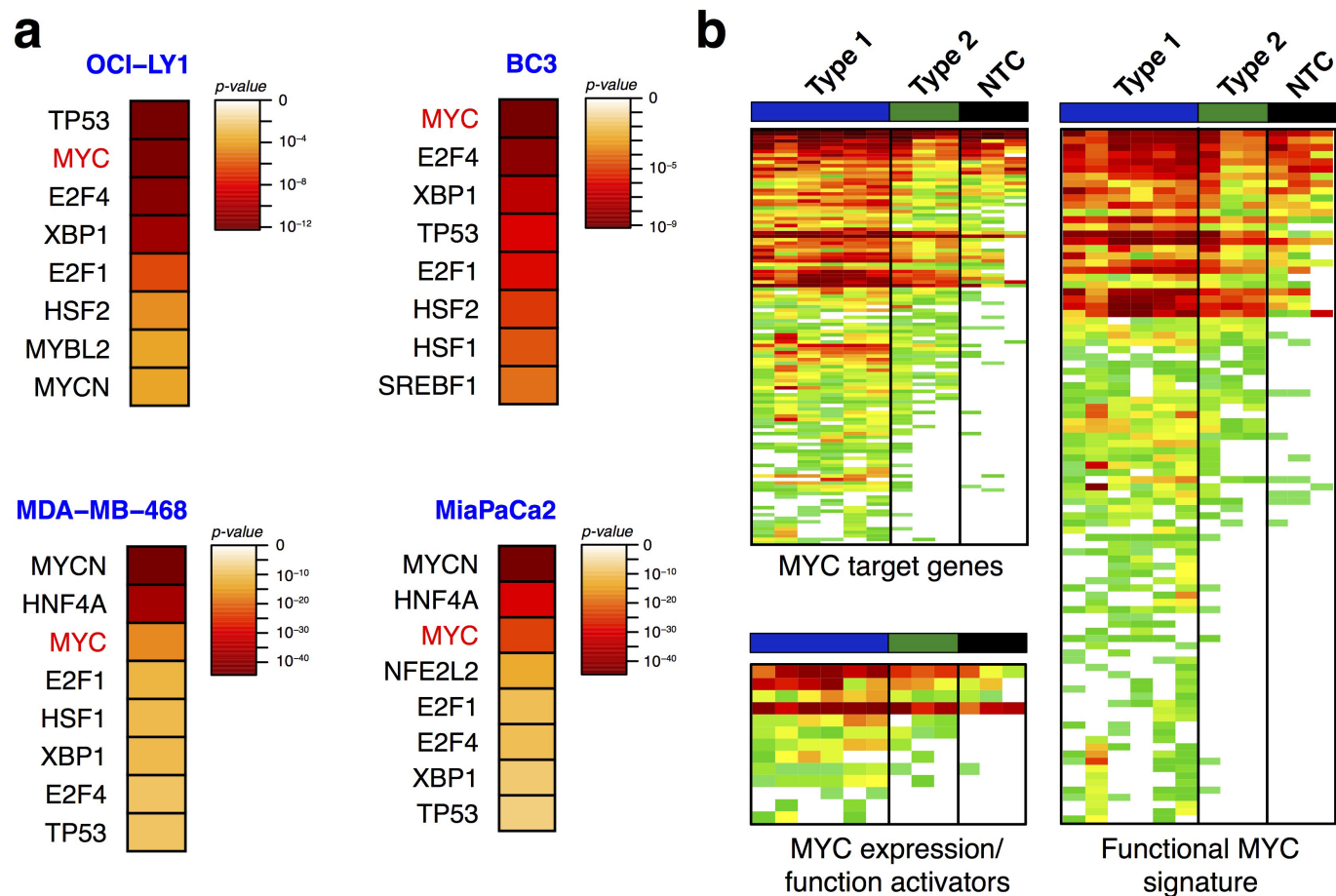
a, Epichaperome levels were altered by AHA1 siRNA knockdown or a control (scramble, Scr) siRNA (left panel) and cell viability, as measured by PARP cleavage, was determined in cells treated for 24 h with increasing concentrations of PU-H71 (0, 0.5, 1 and 2 μ M) (right panel). See also Extended Data Fig. 5g for biochemical signature of cells after AHA1 knockdown. Data are representative of two independent experiments.

b–e, Total protein (**b**, **e**), mRNA (**c**) and multimeric species of indicated chaperome members (**d**) levels were monitored in MDA-MB-468 type 1 cells in which several concentrations of siRNAs against HSP90 α/β ($n = 14$)

were titrated in. 1 and 8 are control scramble; 2, 3, 4, 9, 10, 11 are 0.915, 1.83, 3.66, 0.366, 1.83 and 14.64 nM of siRNA no. 1, respectively; and 5, 6, 7, 12, 13, 14 are 0.366, 0.915, 3.66, 0.0229, 0.0915 and 0.366 nM of siRNA no. 2, respectively. Cell viability in each condition was monitored by LDH release. Values for each experimental condition (as percentage control scramble) were quantified and are noted under the native gels in **a**. HMW, high molecular weight. For gels, experiments were repeated independently twice with representative data shown. For graphs, mean \pm s.d., $n = 6$.

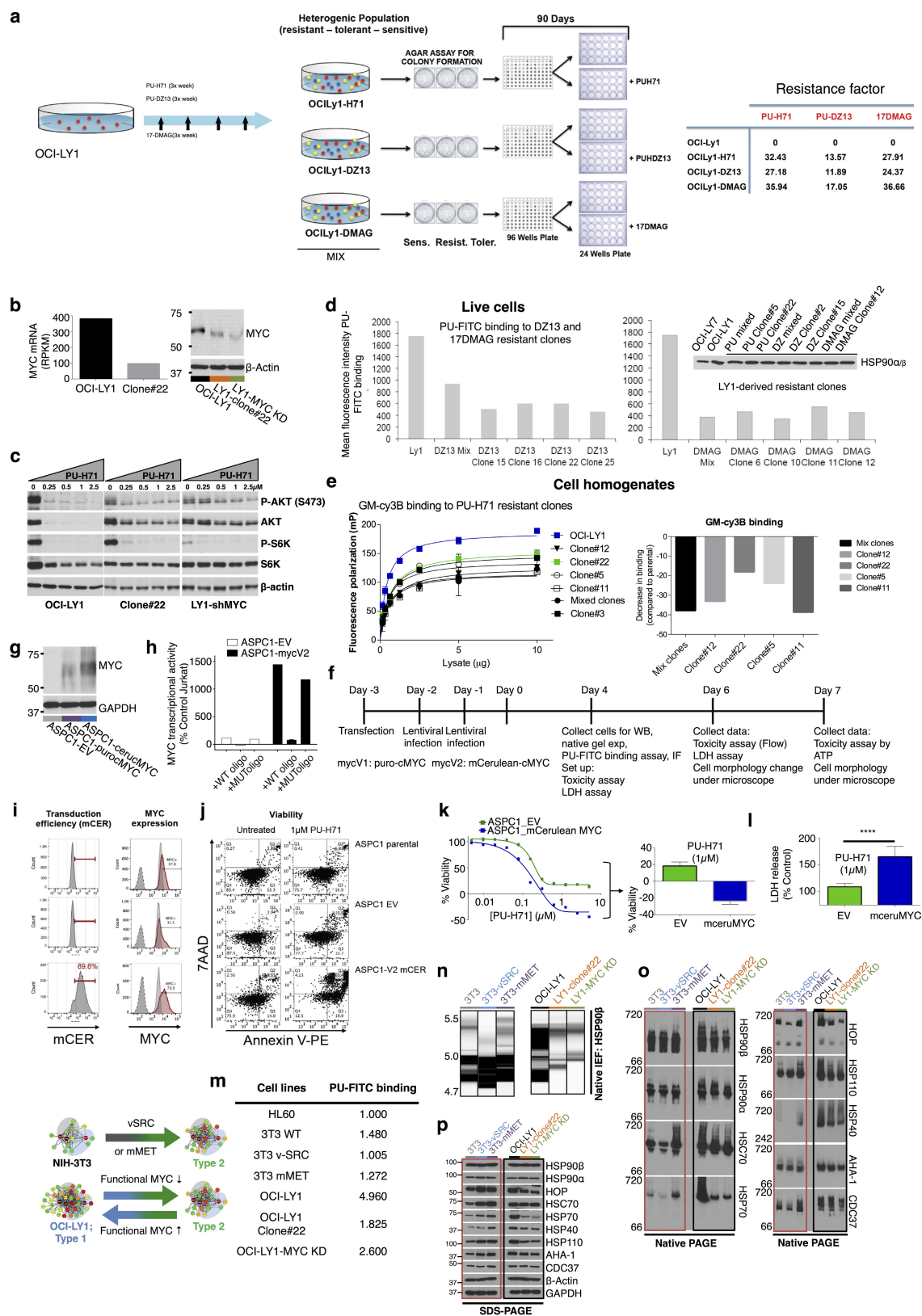
f, Changes in chaperome members (mRNA) were monitored following siRNA knockdown of the indicated individual chaperome members in MDA-MB-468 and ASPC1 cells. Error bars show mean \pm s.d., $n = 6$.

g, Same as for **b–d** in ASPC1, type 2 cells (1 through 4 siRNA concentrations, as in **d**. Error bars represent mean \pm s.d., $n = 6$, unpaired t -test, NS, not significant). For uncropped gel data, see Supplementary Fig. 1.



Extended Data Figure 8 | MYC is a driver of chaperome rewiring into the epichaperome. a, The top eight transcriptional regulators ranked by significant overlap between dataset genes and known targets regulated by the indicated transcription regulators in type 1 tumours. **b,** Heat map

illustrating the enrichment in type 1 tumours of MYC target genes, MYC transcriptional signature and of proteins that regulate MYC function and/or expression.

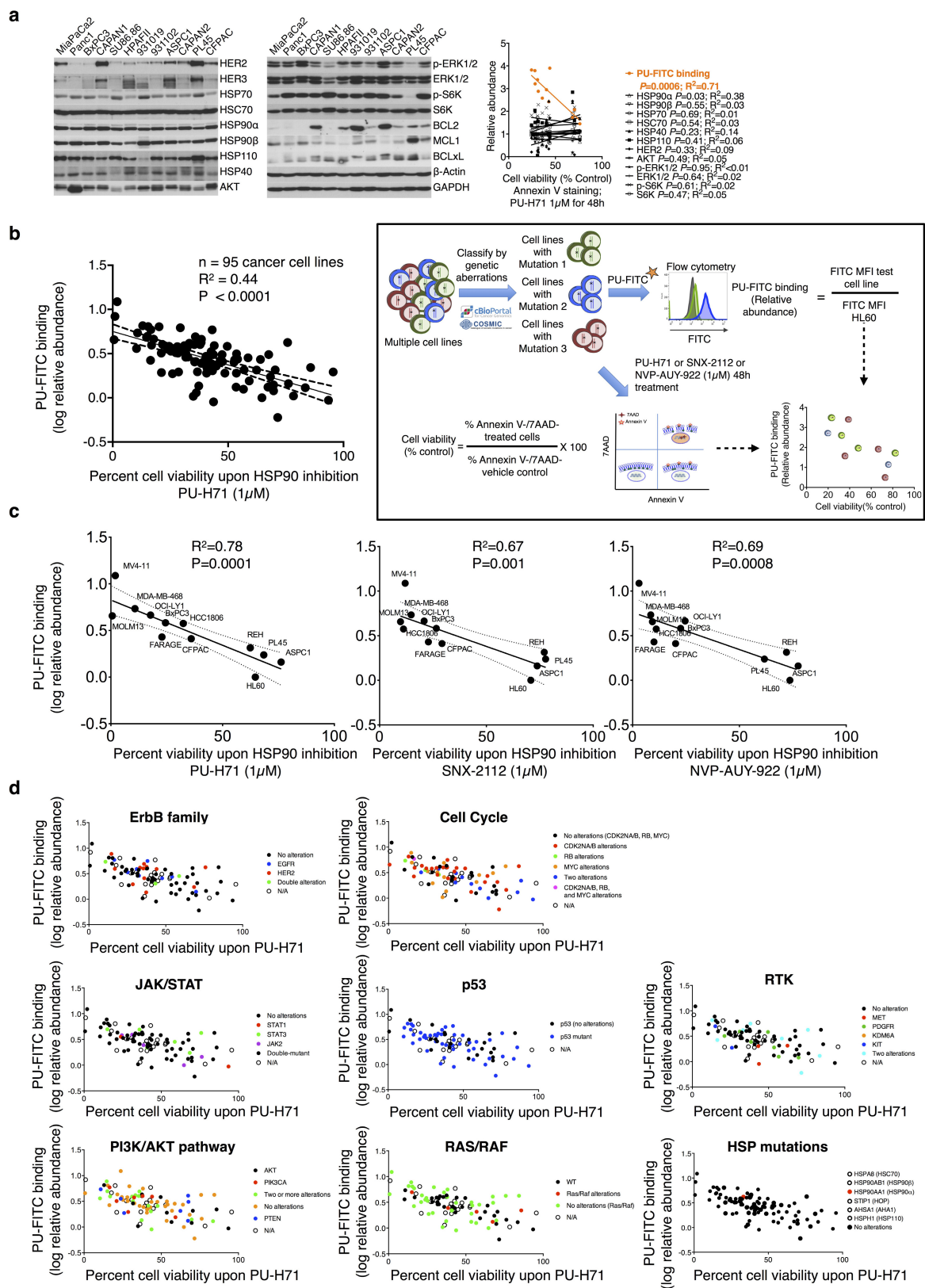


Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Chaperome rewiring into the epichaperome is fuelled by powerful transcription activators such as MYC. **a–e**,

Establishment and characterization of HSP90-inhibitor-resistant cells. **a**, Schematic detailing the establishment and separation of clones cross-resistant to PU-H71, PU-DZ13 and 17-DMAG. These compounds are chemically distinct HSP90 inhibitors. **b**, RNA-seq and western blot analyses of clone no. 22 indicate that cellular resistance to HSP90 inhibitors is associated with MYC downregulation. RPKM, reads per kilobase of exon per million mapped reads. Western blot data are representative of two independent experiments. **c**, HSP90 remains functional in the HSP90-inhibitor resistant clones, as well as in cancer cells in which MYC expression is reduced by shRNA knockdown. Cells were treated for 24 h with the indicated concentrations of PU-H71, and HSP90 client protein function (p-AKT and p-S6K levels) was analysed by western blot. For gels, experiments were repeated independently twice with a representative gel shown. **d**, PU-FITC binding to the indicated resistant clones ($n = 10$) presented as relative, mean fluorescence intensity values, was measured by flow cytometry. The parental OCI-LY1 (type 1 tumour cell) is shown for comparison. Inset shows total HSP90 levels measured by western blot in the indicated clones ($n = 8$). IB; anti-HSP90 (F8) sc-13119. **e**, Binding of a fluorescently labelled geldanamycin derivative (GM-cy3B) to the indicated cell homogenates was measured by fluorescence polarization. Graph shows mean from three technical replicates. **f**, Experimental design for the creation and characterization of

MYC-expressing ASPC1 cells. **g**, Levels of MYC and HSP90 were analysed by western blot in the indicated infection conditions (day 4 post-lentiviral infection). Data were repeated independently twice with a representative blot shown. **h**, Transcriptional activity of MYC in infected cells was measured using the TransAM c-Myc Transcription Factor ELISA. Mean from three technical replicates is shown. **i**, Flow cytometry confirmed the expression of MYC in infected ASPC1 cells. mCerulean and MYC were co-expressed with a 2A peptide linker, which was self-cleaved after protein translation. Data were repeated independently twice with representative data shown. **j–l**, Viability of ASPC1 cells infected with either empty vector or MYC was assessed using an assay that quantifies annexin V/7AAD-stained cells (**j**), ATP levels (**k**), LDH release (**l**) following treatment with PU-H71, as indicated. **k**, Mean of four technical replicates. **l**, Error bars show mean \pm s.d., $n = 6$, unpaired t -test, **** $P < 0.0001$. **m–p**, HSP90 oncogenic kinase clients do not require the epichaperome for cell transforming activity. **m**, PU-FITC binding to the indicated live cells presented as a ratio (fluorescent signal of measured cells over signal in HL60 cells; HL60, internal standard). **n–p**, Changes in multimeric chaperome complexes (**n**, **o**) and total protein (**p**) in the indicated conditions. All data were repeated independently twice with representative data shown. OCI-LY1 (type 1 cells) and OCI-LY1 rewired to type 2 following MYC loss are presented for direct comparison purposes. For uncropped gels, see Supplementary Fig. 1.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | The apoptotic response profile of a panel of cancer cells following HSP90 inhibition is independent of levels of chaperome members, HSP90 client proteins and anti-apoptotic molecules, tissue of origin or causal genetic mutations. **a**, Total levels of the indicated chaperome members, HSP90 client proteins and anti-apoptotic molecules were analysed by western blot in a panel of pancreatic cancer cells ($n = 12$). GAPDH and β -actin, protein loading controls. Protein levels were quantified and graphed against the viability of these cells upon HSP90 inhibition. A correlative analysis was performed (Pearson's r , two-tailed). Each data point represents a cell line. PU-FITC binding is shown for comparison. **b**, Correlative analysis of epichaperome abundance, as measured by PU-FITC staining, and cell viability upon a 48 h treatment with PU-H71 (1 μ M), as measured by annexin V staining (Pearson's r , two-tailed). Each data point represents a cell line

($n = 95$); data are the mean from two or three biological replicates. Cells representing pancreatic, gastric, lung, and breast cancers, along with lymphomas and leukaemias were chosen for analysis. **c**, Same as above for the treatment of cancer cells ($n = 12$) with three chemically distinct HSP90 inhibitors. **d**, Same as **b**, but for each cell line, known genetic lesions were added. No specific genetic alteration could be found distinguishing the two tumour types; whereas p53-, Ras-, Myc-, HER-, PI3K/AKT-, and JAK- cell cycle-related defects were found in tumours that were sensitive to PU-H71, they were also evident in PU-H71 resistant cells. We found genetic defects in major chaperome members to be rare, with BCP-1 cells only carrying an *HSP90AA1* missense mutation (P596S). No mutations in *HSP90AB1*, *HSPH1*, *HSPA8*, *STIP1* and *AHSA1* were reported in this large panel of cell lines. These were obtained from the cBioPortal for Cancer Genomics website (<http://www.cbioportal.org/>).

Molecular basis of Lys11–polyubiquitin specificity in the deubiquitinase Cezanne

Tycho E. T. Mevissen¹, Yogesh Kulathu^{1*†}, Monique P.C. Mulder^{2*†}, Paul P. Geurink^{2†}, Sarah L. Maslen¹, Malte Gersch¹, Paul R. Elliott¹, John E. Burke^{1†}, Bianca D.M. van Tol², Masato Akutsu^{1†}, Farid El Oualid^{2†}, Masato Kawasaki³, Stefan M.V. Freund¹, Huib Ovaa^{2†} & David Komander¹

The post-translational modification of proteins with polyubiquitin regulates virtually all aspects of cell biology. Eight distinct chain linkage types co-exist in polyubiquitin and are independently regulated in cells. This ‘ubiquitin code’ determines the fate of the modified protein¹. Deubiquitinating enzymes of the ovarian tumour (OTU) family regulate cellular signalling by targeting distinct linkage types within polyubiquitin², and understanding their mechanisms of linkage specificity gives fundamental insights into the ubiquitin system. Here we reveal how the deubiquitinase Cezanne (also known as OTUD7B) specifically targets Lys11-linked polyubiquitin. Crystal structures of Cezanne alone and in complex with monoubiquitin and Lys11-linked diubiquitin, in combination with hydrogen–deuterium exchange mass spectrometry, enable us to reconstruct the enzymatic cycle in great detail. An intricate mechanism of ubiquitin-assisted conformational changes activates the enzyme, and while all chain types interact with the enzymatic S1 site, only Lys11-linked chains can bind productively across the active site and stimulate catalytic turnover. Our work highlights the plasticity of deubiquitinases and indicates that new conformational states can occur when a true substrate, such as diubiquitin, is bound at the active site.

The 16 human OTU family deubiquitinases (DUBs) are key regulators of the ubiquitin code. Small OTU DUBs of the OTUD and OTUB subfamilies, which have catalytic cores of approximately 130–220 residues, employ distinct mechanisms to achieve linkage specificity^{2–5}, but the physiological roles of most of these DUBs are unclear. By contrast, the A20-like OTU subfamily, identified by a larger catalytic domain (300–350 residues, Fig. 1a), has been well studied. A20 is a tumour suppressor and negative feedback regulator of NF- κ B signalling^{6,7}; TRABID (also known as ZRANB1) regulates transcription^{8,9} by targeting Lys29- and Lys33-linked chains¹⁰; and VCIPI is associated with valosin-containing protein (VCP, also known as p97; ref. 11) and cleaves Lys48 and Lys11 linkages².

Cezanne regulates inflammation and NF- κ B signalling^{12–14}, T-cell activation¹⁵, epidermal growth factor receptor (EGFR) trafficking¹⁶, and homeostasis of the transcription factors HIF-1 α and HIF-2 α ^{17,18}. Cezanne and Cezanne2 (also known as OTUD7A; 61% identity with Cezanne) (Fig. 1a) are the only DUBs known to be specific for Lys11-linked polyubiquitin^{2,17,19}. Lys11 specificity is encoded in the catalytic domain of Cezanne^{2,19} (Fig. 1b), and extends to Lys11 linkages within Lys11/Lys63- and Lys11/Lys48-branched chains (Extended Data Fig. 1a). A fluorescence resonance energy transfer (FRET)-based kinetic cleavage assay²⁰ showed that Cezanne has similar Michaelis constant (K_M) values for Lys11-, Lys63- and Lys48-linked diubiquitin, but a significantly higher

catalytic turnover number (k_{cat}) for Lys11 diubiquitin (Fig. 1c and Extended Data Fig. 1b, c).

To understand the k_{cat} -driven specificity of Cezanne, we determined crystal structures of Cezanne alone (Cez apo, 2.2 Å), in complex with monoubiquitin (Cez–Ub, 2.0 Å; two distinct complexes in the asymmetric unit), and bound to Lys11 diubiquitin (Cez–Lys11 diUb, 2.8 Å) (Fig. 1d, Extended Data Figs 2, 3, Extended Data Table 1 and Methods). Determination of the structure of Cez–Lys11 diUb used covalent diubiquitin activity-based probes (ABPs)²¹ (Fig. 1e, f). In addition, an A20–Ub complex was determined at 2.85 Å (Fig. 1g, Extended Data Table 1 and Methods). The structures of Cezanne resemble those of A20^{22,23} and TRABID¹⁰ with some topological differences (Extended Data Fig. 2).

Monoubiquitin-bound structures have not previously been available for the A20-like subfamily, and the Cez–Ub and A20–Ub complexes reveal a conserved S1 ubiquitin-binding site that is distinct from other OTU DUBs (Extended Data Fig. 3). The apo states of A20, Cezanne and TRABID feature an unobstructed S1 site, and A20–Ub is highly similar to unliganded A20 (r.m.s.d. 0.54 Å; Extended Data Fig. 3d).

Monoubiquitin complexes depict product-bound rather than substrate-bound forms, and do not explain the specificity of DUBs. The Cez–Lys11 diUb complex represents a substrate-bound state that reveals a new S1' site, and together the structures explain the specificity of Cezanne. Indeed, large-scale conformational changes between individual structures, in which the S1' site is transiently formed and lost, delineate the catalytic cycle for Lys11 diubiquitin hydrolysis in Cezanne (Fig. 2, Extended Data Fig. 4 and Supplementary Videos 1, 2).

Cez apo is autoinhibited owing to the conformation of the Cys-loop² (residues 187–193) that precedes catalytic Cys194. Asn193 occupies the channel that binds the C-terminal tail of the distal ubiquitin, and the catalytic His358 is unable to deprotonate Cys194 (Fig. 2b and Extended Data Fig. 4). A key structural residue, His197, stabilizes the autoinhibited Cys-loop conformation (Fig. 2b).

Substantial conformational changes take place upon substrate binding (Fig. 2a). The distal ubiquitin binds the accessible S1 site, while the proximal ubiquitin interacts with a new S1' site formed by the α 3– α 4 linker (S1'-loop hereafter) and by helices α 1 and α 2, which change in position and length (Fig. 2a and Extended Data Fig. 4). These rearrangements enable hydrophobic residues (Leu155, Met203 and Phe206) to come together and bind the proximal ubiquitin (Extended Data Fig. 4). His197 no longer coordinates the Cys-loop but now binds the S1'-loop (Fig. 2c). As a result, the Cys-loop moves, forming the oxyanion hole and enabling Cys194 to form the tetrahedral intermediate mimic with the diubiquitin ABP (Extended Data Fig. 2d). Notably, the

¹Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK. ²Division of Cell Biology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ³Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan. [†]Present addresses: Medical Research Council Protein Phosphorylation and Ubiquitylation Unit, Dow Street, Dundee, DD1 5EH, UK (Y.K.); Chemical Immunology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, The Netherlands (M.P.C.M., P.P.G., H.O.); Department of Biochemistry and Microbiology, University of Victoria, 270 Petch Hall, Victoria BC, Canada (J.E.B.); Buchmann Institute for Molecular Life Sciences, Goethe University, 60438 Frankfurt am Main, Germany (M.A.); UbiQ Bio BV, Science Park 408, 1098 XH, Amsterdam, The Netherlands (F.E.O.).

*These authors contributed equally to this work.

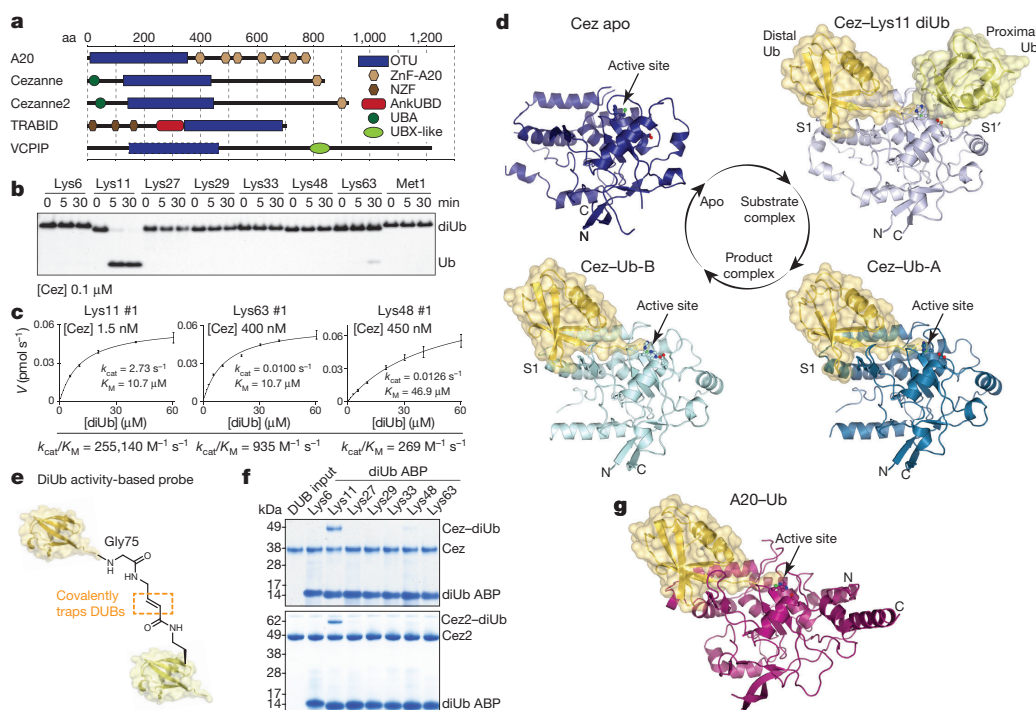


Figure 1 | Cezanne biochemistry and structures. **a**, Domain architecture of A20-like OTU DUBs. **b**, Specificity analysis of the Cezanne OTU domain (residues 129–438). This experiment was performed three times. Ub, ubiquitin; diUb, diubiquitin. **c**, Representative graphs of initial rates and kinetic parameters for Lys11, Lys63 and Lys48 diubiquitin hydrolysis by Cezanne. Assays were performed in triplicate and in at least three independent experiments (Extended Data Fig. 1c). Error bars represent s.d. from the mean. **d**, Cezanne (residues 129–438) structures determined in this study: Cez apo, Cez-Lys11 diUb and Cez-Ub ('QPG'; see Methods).

catalytic His358 does not coordinate Cys194 but remains in an inactive conformation, as the proximal ubiquitin pushes Thr188 of the Cys-loop into the position of His358 in the catalytic centre ('His358-out' conformation; Fig. 2c).

The next step in the cycle is illuminated by the two Cez-Ub complexes, which no longer feature an S1' ubiquitin-binding site (Fig. 2a and Extended Data Fig. 4). Consistently, Cez-Ub shows no interaction with ubiquitin in NMR or fluorescence polarization measurements (Extended Data Fig. 5a, b). Importantly, the two Cez-Ub complexes have different catalytic centres. Cez-Ub-A features the inactive 'His358-out' conformation (Fig. 2d), whereas Cez-Ub-B displays a catalytically competent state, in which Thr188 has moved out and His358 coordinates Cys194 (Fig. 2e and Extended Data Fig. 4). His197 is again a key residue that now stabilizes the Cys-loop in the active state (Fig. 2e), which hydrolyses the acyl intermediate to regenerate Cez apo (Fig. 2a). The active state might also depict the initial Cys deprotonation stage before tetrahedral intermediate formation (within transition I, Fig. 2a). Indeed, Cys194 and His358 are essential for hydrolysis of diubiquitin and a monoubiquitinated fluorescent Lys-Gly (KG) dipeptide (Ub-KG²⁴; Extended Data Fig. 5c, d).

Structure-based mutagenesis confirmed key mechanistic features of Cezanne. Mutation of Asn193 to Leu or Met stabilized autoinhibition of Cez apo by improving contacts in the ubiquitin-binding channel and abrogated DUB activity and ubiquitin binding (Fig. 3a, b and Extended Data Fig. 5e, f). Mutation of His197 or its coordinating residue Asp210 to Ala reduced cleavage of diubiquitin and Ub-KG* (Extended Data Fig. 5g, h). Consistent with a key structural rather than catalytic role for His197, Cezanne H197A showed residual activity, and mutations maintaining its coordinating capabilities were only mildly impaired (Extended Data Fig. 5i–k). Notably, mutation of the corresponding residue in A20, His106, also reduced A20 DUB activity (Extended Data Fig. 5l).

Both Cez-Ub complexes in the asymmetric unit are depicted. The OTU domain is shown in cartoon representation with active site residues highlighted, and ubiquitin moieties are shown under transparent surfaces. **e**, Schematic of diubiquitin ABP. **f**, Probe assay of Cezanne (residues 129–438, top) and Cezanne2 (residues 1–462, bottom) with differently linked diubiquitin ABPs. Experiments were replicated twice. **g**, Crystal structure of A20-Ub (residues 1–366, see Methods). For gel source images, see Supplementary Fig. 1.

Finally, hydrogen-deuterium exchange mass spectrometry (HDX-MS) confirmed the 'footprints' of bound ubiquitin moieties, and the conformational transitions observed crystallographically (Extended Data Figs 6, 7). The elongated α2-helix that is disrupted in complex structures (Fig. 2a) displayed an uncharacteristically high hydrogen-deuterium exchange for helices in Cez apo, suggesting that this helix may be dynamic in solution (Extended Data Fig. 6a, b). Mutations that destabilize helix α2 (I156G or L155G/I156G) impaired Cezanne activity (Fig. 3c, d and Extended Data Figs 5e, 6c), indicating that there is a required order-disorder transition in this region. Moreover, despite the lack of direct contacts, these mutants reduced ubiquitin binding to the S1 site (Extended Data Fig. 5f). Ubiquitin binding substantially improved the thermal stability of Cezanne (Extended Data Fig. 8a), and increased hydrogen-deuterium exchange was observed in multiple elements corresponding to the S1' site upon ubiquitin release (Extended Data Fig. 7a, d). This reveals that S1 site ubiquitin binding is coupled to S1' site dynamics, and primes the enzyme for substrate discrimination and catalysis. Together, mutagenesis and HDX-MS confirm the observed conformational transitions in the catalytic cycle (Fig. 2).

We next studied how these transitions impose linkage specificity. The small differences in K_M among Lys11-, Lys63- and Lys48-linked diubiquitin (Fig. 1c) suggest that substrate engagement is primarily driven by the exposed S1 site, which involves the Cezanne α5–α6 helical arm and a hydrophobic loop (Fig. 3e). These elements contact both Ile44 and Ile36 patches of ubiquitin as well as its C terminus. Mutations in the helical arm or in the channel for the ubiquitin C terminus abrogated hydrolysis of Lys11 diubiquitin and Ub-KG* (Fig. 3f and Extended Data Fig. 8b) and reactivity towards the Lys11 diubiquitin ABP (Fig. 3g). Monoubiquitin-binding assays indicated a strong interface (dissociation constant (K_D) 9.3 μM for wild-type Cezanne,

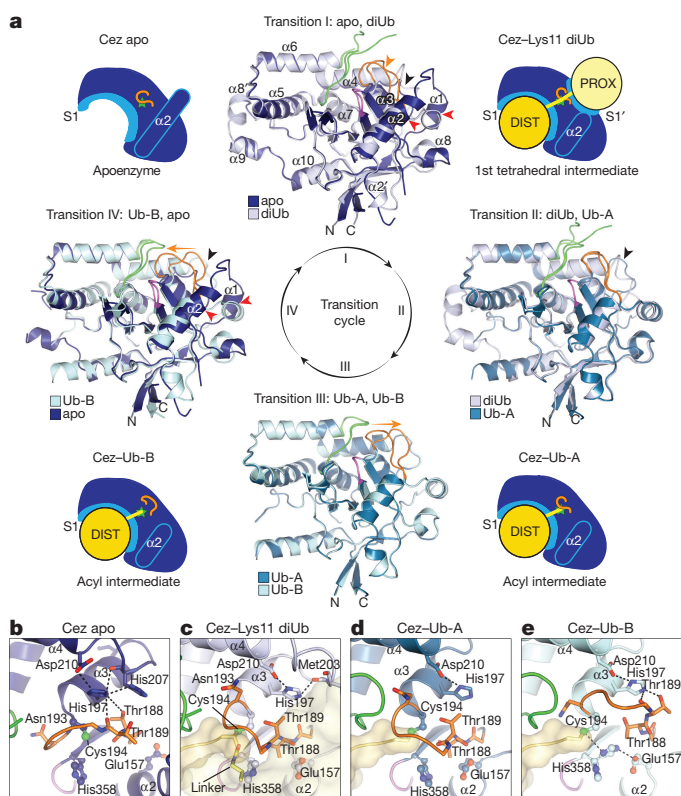


Figure 2 | Conformational changes in the Cezanne catalytic cycle. **a**, Schematic cartoons of the determined structures (Fig. 1d) highlight four catalytic states of the reaction cycle (green star, active site; orange line, Cys-loop). In between these states, superpositions of the OTU domain show transitions I–IV. Loops are coloured orange (Cys-loop), green (V-loop) and purple (His-loop). Transition I (diubiquitin substrate binding) is characterized by conformational changes around the catalytic centre, including the Cys-loop (orange arrowhead), α1 and α2 helices (red arrowheads) and the S1'-loop (black arrowhead). In transition II (proximal ubiquitin release), a second S1'-loop rearrangement relocates S1' site residues (black arrowhead). Transition III features a Cys-loop movement. Several structural changes regenerate Cezanne apo in transition IV. Also see Supplementary Video 1. **b–e**, Active site close-up views of the four states. Selected residues are shown as sticks. Hydrogen bond networks of His197 and the catalytic centre are indicated. Also see Extended Data Fig. 4 and Supplementary Video 2.

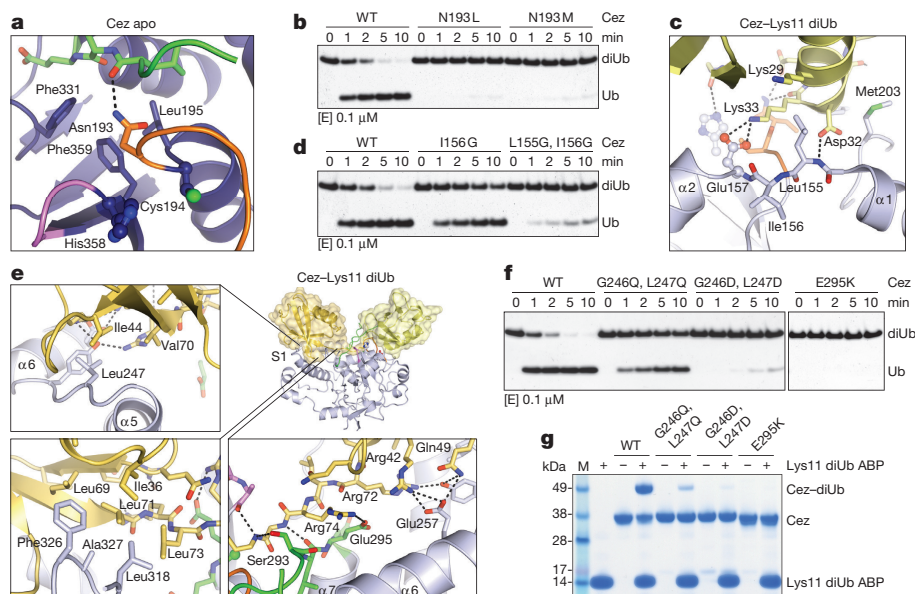


Figure 3 | Mutational analysis of Cezanne dynamics. **a**, In Cez apo, Asn193 of the Cys-loop (orange) blocks the distal ubiquitin-binding channel and is positioned above a hydrophobic pocket. **b**, DUB assay of Asn193 mutants that 'lock' the enzyme in the apo state. **c**, The α1–α2 linker in Cez-Lys11 diUb contains residues Leu155 and Ile156 adjacent to Glu157. **d**, DUB assay of mutants with a destabilized α2-helix (see Extended Data Fig. 6c). **e**, The S1 site of Cezanne contacts the Ile44 patch (top left), the Ile36 patch (bottom left) and the C-terminal tail of the distal ubiquitin (bottom right). **f**, DUB assay of S1 site mutants. WT, wild-type. **g**, Lys11 diubiquitin ABP probe assay of S1 site mutants. All assays were performed at least twice with consistent results. For gel source images, see Supplementary Fig. 1.

0.43 μM for Cezanne C194A; Extended Data Fig. 8c, d). Moreover, monoubiquitin or differently linked diubiquitin bound Cezanne C194A similarly in pull-down assays, and this depended on a free ubiquitin C terminus; once the C terminus is removed, interactions are lost except for Lys11 diubiquitin, as this chain type can bind across the active site (Extended Data Fig. 8e). Hence, the S1 site is responsible for substrate recruitment.

DUB linkage specificity relies on careful positioning of the proximal ubiquitin, which interacts with Cezanne via an unusual surface, involving Thr12, Glu16 and its α-helix (Asp32, Lys33, Glu34, Gly35; Fig. 3c and Extended Data Fig. 9a). A similar interface is also used by UBE2S, the Lys11-specific E2 enzyme²⁵. Proximal ubiquitin engagement transiently forms the S1' site, enabling catalysis. However, consistent with the weak interface, mutations in hydrophobic S1' site residues had little effect on DUB activity or probe reactivity (Extended Data Fig. 9b, c).

Importantly, a direct interaction between Lys33 of the proximal ubiquitin and Glu157 of the catalytic centre (Fig. 3c) affects Cezanne catalytic turnover. Cezanne hydrolysed Lys11 diubiquitin with proximal K33A or K33E mutations less efficiently than wild-type Lys11 diubiquitin (Extended Data Fig. 9d). Cezanne E157K cleaved Ub-KG* (Fig. 4a) but showed severely impaired activity towards Lys11 diubiquitin substrates (Fig. 4b and Extended Data Fig. 9e). Kinetic analyses of this reaction reveal a drop in k_{cat} for Lys11 diubiquitin hydrolysis by Cezanne E157K as compared to wild-type Cezanne, whereas Lys63 and Lys48 cleavage was almost unaffected (Fig. 4c and Extended Data Fig. 1c, d). Qualitative assays confirm that Cezanne E157K is less specific than wild-type Cezanne (Extended Data Fig. 9f, g). Lys33 binding to Glu157 requires Cezanne to be in its inactive 'His358-out' conformation. Hence, Glu157 is important for substrate selection but not for catalysis, and while it can coordinate His358, its presence is not essential for the catalytic dyad of Cezanne.

Together, our findings illuminate a complete DUB catalytic cycle and reveal the molecular basis of Cezanne's Lys11 specificity (Fig. 4d). The dynamic Cez apo state engages polyubiquitin at the S1 site, releasing Cys-loop-mediated autoinhibition. This primes the enzyme for substrate discrimination by S1' site remodelling. Only Lys11-linked polyubiquitin forms favourable contacts with the S1' site, which opens transiently to enable interactions between Lys33 and Glu157, and formation of this contact improves catalytic turnover. After isopeptide bond hydrolysis, conformational changes destroy the S1' site and expel the proximal ubiquitin. The remaining product-bound monoubiquitin complex is resolved by further rearrangements that align the catalytic

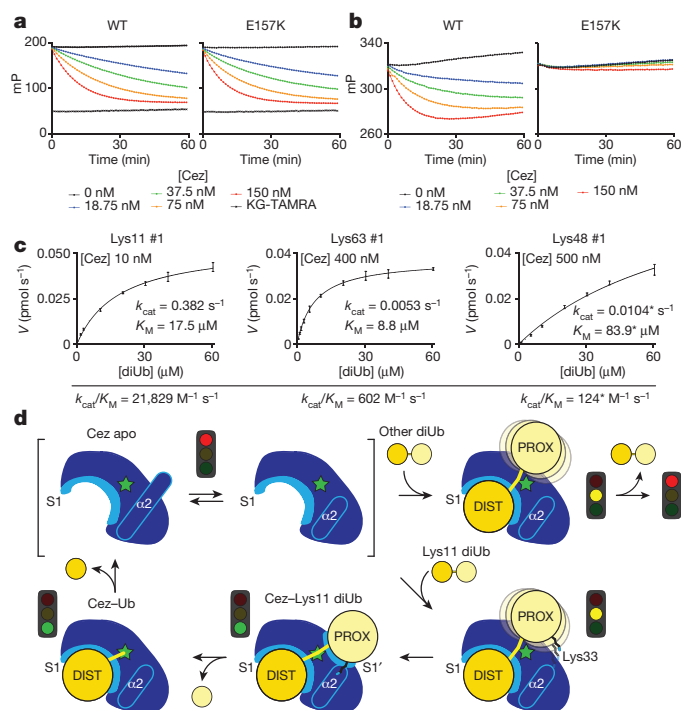


Figure 4 | Basis of Lys11 specificity and model of Cezanne mechanism.

a, b, Fluorescence polarization cleavage assays comparing wild-type and E157K Cezanne using Ub-KG* (**a**) and FIAH-tagged Lys11-linked diubiquitin (**b**). mP, millipolarization unit. Fluorescence polarization measurements were performed in triplicate in at least two independent experiments. **c**, Summary of Cezanne E157K diubiquitin cleavage kinetics. Compared to wild-type Cezanne (Fig. 1c), this mutant is impaired in cleaving Lys11 linkages. Assays were performed in triplicate and in at least two independent experiments. *These values suffer from technical limitations (for more detail, see Extended Data Fig. 1d). Error bars represent s.d. from the mean. **d**, Model of Cezanne mechanism. The apo-enzyme is autoinhibited but dynamic, and recruits a substrate with its accessible S1 site. Only Lys11-linked diubiquitin can interact specifically with the formed S1' site, involving an activating interaction between Cezanne and ubiquitin Lys33. After cleavage, the S1' site is lost and the proximal moiety expelled. Subsequent hydrolysis and distal ubiquitin release recreates the Cezanne apo state.

centre to hydrolyse the acyl intermediate. Upon distal ubiquitin release, Cezanne is restored to its autoinhibited state (Fig. 4d).

Our work highlights the potential plasticity of DUBs, which, in the case of Cezanne, results in marked conformational transitions along the reaction cycle (Supplementary Videos 1, 2). With the rising importance of DUBs as drug targets, insights into conformational flexibility are essential. Indeed, small molecule DUB inhibitors targeting Cezanne or TRABID, which also requires conformational domain rearrangements¹⁰, may open new avenues for the treatment of cancer and inflammation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April; accepted 8 September 2016.

Published online 12 October 2016.

- Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).
- Mevisen, T. E. T. et al. OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell* **154**, 169–184 (2013).
- Keusekotten, K. et al. OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. *Cell* **153**, 1312–1326 (2013).
- Wiener, R., Zhang, X., Wang, T. & Wolberger, C. The mechanism of OTUB1-mediated inhibition of ubiquitination. *Nature* **483**, 618–622 (2012).
- Juang, Y.-C. et al. OTUB1 co-opts Lys48-linked ubiquitin recognition to suppress E2 enzyme function. *Mol. Cell* **45**, 384–397 (2012).

- Hymowitz, S. G. & Wertz, I. E. A20: from ubiquitin editing to tumour suppression. *Nat. Rev. Cancer* **10**, 332–341 (2010).
- Wertz, I. E. et al. Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. *Nature* **528**, 370–375 (2015).
- Tran, H., Hamada, F., Schwarz-Romond, T. & Bienz, M. Trabid, a new positive regulator of Wnt-induced transcription with preference for binding and cleaving K63-linked ubiquitin chains. *Genes Dev.* **22**, 528–542 (2008).
- Jin, J. et al. Epigenetic regulation of the expression of I12 and I23 and autoimmune inflammation by the deubiquitinase Trabid. *Nat. Immunol.* **17**, 259–268 (2016).
- Licchesi, J. D. F. et al. An ankyrin-repeat ubiquitin-binding domain determines TRABID's specificity for atypical ubiquitin chains. *Nat. Struct. Mol. Biol.* **19**, 62–71 (2011).
- Christianson, J. C. & Ye, Y. Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nat. Struct. Mol. Biol.* **21**, 325–335 (2014).
- Hu, H. et al. OTUD7B controls non-canonical NF- κ B activation through deubiquitination of TRAF3. *Nature* **494**, 371–374 (2013).
- Enesa, K. et al. NF- κ B suppression by the deubiquitinating enzyme Cezanne: a novel negative feedback loop in pro-inflammatory signaling. *J. Biol. Chem.* **283**, 7036–7045 (2008).
- Luong, A. et al. Cezanne regulates inflammatory responses to hypoxia in endothelial cells by targeting TRAF6 for deubiquitination. *Circ. Res.* **112**, 1583–1591 (2013).
- Hu, H. et al. Otud7b facilitates T cell activation and inflammatory responses by regulating Zap70 ubiquitination. *J. Exp. Med.* **213**, 399–414 (2016).
- Pareja, F. et al. Deubiquitination of EGFR by Cezanne-1 contributes to cancer progression. *Oncogene* **31**, 4599–4608 (2012).
- Bremm, A., Moniz, S., Mader, J., Rocha, S. & Komander, D. Cezanne (OTUD7B) regulates HIF-1 α homeostasis in a proteasome-independent manner. *EMBO Rep.* **15**, 1268–1277 (2014).
- Moniz, S. et al. Cezanne regulates E2F1-dependent HIF2 α expression. *J. Cell Sci.* **128**, 3082–3093 (2015).
- Bremm, A., Freund, S. M. V. & Komander, D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nat. Struct. Mol. Biol.* **17**, 939–947 (2010).
- Geurink, P. P. et al. Development of diubiquitin-based FRET probes to quantify ubiquitin linkage specificity of deubiquitinating enzymes. *ChemBioChem* **17**, 816–820 (2016).
- Mulder, M. P. C., El Oualid, F., ter Beek, J. & Ova, H. A native chemical ligation handle that enables the synthesis of advanced activity-based probes: diubiquitin as a case study. *ChemBioChem* **15**, 946–949 (2014).
- Komander, D. & Barford, D. Structure of the A20 OTU domain and mechanistic insights into deubiquitination. *Biochem. J.* **409**, 77–85 (2008).
- Lin, S.-C. et al. Molecular basis for the unique deubiquitinating activity of the NF- κ B inhibitor A20. *J. Mol. Biol.* **376**, 526–540 (2008).
- Geurink, P. P., El Oualid, F., Jonker, A., Hameed, D. S. & Ova, H. A general chemical ligation approach towards isopeptide-linked ubiquitin and ubiquitin-like assay reagents. *ChemBioChem* **13**, 293–297 (2012).
- Wickliffe, K. E., Lorenz, S., Wemmer, D. E., Kuriyan, J. & Rape, M. The mechanism of linkage-specific ubiquitin chain elongation by a single-subunit E2. *Cell* **144**, 769–781 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Wakatsuki and N. Matsugaki for access to KEK PF BL-1A, and beam-line scientists at ESRF ID23-1, ID29 and Diamond I02 and I03; R. Williams and M. Skehel for help and discussion on HDX-MS; and C. Johnson and S. McLaughlin for help with biophysics. Access to DLS was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (contract no. 283570). This work was supported by the Medical Research Council (U105192732 (D.K.)), the European Research Council (309756 (D.K.); 281699 (H.O.)), the Lister Institute for Preventive Medicine (D.K.), the British Heart Foundation (PG11/109/29247 (J.E.B.)), and a Netherlands Organization for Scientific Research VICI grant (724.013.002 (H.O.)). T.E.T.M. was supported by the Marie Curie ITN UPStream and Y.K. by Marie Curie and EMBO Long Term Fellowships.

Author Contributions D.K. directed the research. T.E.T.M. performed all biochemical experiments, crystallized and determined the structure of the Cez-Lys11 diUb complex, and refined and analysed all structures. Y.K. crystallized and determined the structure of Cez-Ub and A20-Ub and performed preliminary biochemistry. M.P.C.M. and F.E.O. designed and generated diubiquitin ABPs. P.P.G. designed and generated (with B.D.M.v.T.) diubiquitin-based FRET probes. S.L.M. performed HDX-MS. J.E.B. performed preliminary HDX-MS. M.G. performed TSA assays. P.R.E., M.A. and M.K. helped with data collection and structure determination. S.M.V.F. performed NMR analysis. H.O. guided chemical biology efforts. T.E.T.M. and D.K. analysed the data and wrote the manuscript with help from all authors.

Author Information Coordinates and structure factors have been deposited with the Protein Data Bank under accession codes 5LRU (Cez apo), 5LRV (Cez-Lys11 diUb), 5LRW (Cez-Ub) and 5LRX (A20-Ub). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.K. (dk@mrc-lmb.cam.ac.uk).

METHODS

Cloning and site-directed mutagenesis. A codon-optimized human Cezanne gene (*Otud7b*) for bacterial expression was obtained from GeneArt Gene Synthesis (Invitrogen) and cloned into pOPIN vectors²⁶ using the In-Fusion HD Cloning Kit (Clontech) according to the manufacturer's instructions. Site-directed mutagenesis of Cezanne, A20 and ubiquitin (Ub) was performed using the QuikChange method.

Protein expression and purification. Proteins were expressed in *E. coli* Rosetta2 (DE3) pLacI (Novagen) from the pOPIN-E (Cezanne) or pGEX6P1 (A20) vectors. The pOPIN-E vector introduces a C-terminal His6 tag and pGEX6P1 features a PreScission protease-cleavable N-terminal GST tag. Cells were grown at 37 °C to an optical density (OD)₆₀₀ of 0.8–1.0 and induced with 0.2–0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 18–20 h at 18 °C. Expression was performed in 2–4 L 2× TY medium supplemented with appropriate antibiotics.

All purification steps were performed at 4 °C. Bacterial cells were resuspended in 40–80 ml lysis buffer (25 mM Tris (pH 8.5), 200 mM NaCl, 5 mM DTT (A20) or 2 mM β-mercaptoethanol (Cezanne), 1 mg/ml lysozyme, 0.1 mg/ml DNaseI, one EDTA-free Complete Protease Inhibitor Cocktail tablet), sonicated and cleared by centrifugation at 20,000g for 35 min. His-tagged constructs were affinity purified with TALON Superflow resin (GE Healthcare). Resin (1–2 ml) was incubated with the cleared lysate for 5–10 min and washed with 1 l buffer A (25 mM Tris (pH 8.5), 200 mM NaCl, 2 mM β-mercaptoethanol). Protein was eluted with 5–10 ml buffer A supplemented with 250 mM imidazole and subsequently dialysed against buffer B (25 mM Tris (pH 8.5), 5 mM DTT) plus 50 mM NaCl before further purification (see below). Cleared lysates of GST fusion proteins were incubated with 2–4 ml glutathione sepharose 4B resin (GE Healthcare) for 1 h under constant agitation. The beads were washed with 2 l buffer B supplemented with 500 mM NaCl and 0.5 l buffer B plus 50 mM NaCl. GST tag cleavage was performed overnight with 50–100 μg GST-tagged PreScission protease on the resin. Protein was eluted with buffer B plus 50 mM NaCl before further purification.

All proteins were subjected to anion-exchange chromatography (Resource Q, GE Healthcare) in buffer B with a salt gradient of 50–500 mM NaCl, and subsequent size-exclusion chromatography (HiLoad 16/60 Superdex 75, GE Healthcare) in buffer B supplemented with 200 mM NaCl. Peak fractions were pooled, concentrated to 2–25 mg/ml using Amicon 10 kDa MWCO Ultra Centrifugal Filters (Millipore), frozen in liquid nitrogen and stored at –80 °C.

Qualitative DUB assays. Qualitative *in vitro* deubiquitination assays were performed as described²⁷. In short, DUBs were diluted in DUB dilution buffer (25 mM Tris (pH 7.5), 150 mM NaCl, 10 mM DTT) to 2× indicated concentrations and pre-incubated for 10 min at room temperature. 10 μM stocks (2× final concentration) of differently linked diUb substrates were prepared in 2× DUB reaction buffer (100 mM Tris (pH 7.5), 100 mM NaCl, 10 mM DTT). To start the hydrolysis reaction, DUB and substrate solutions were mixed in a 1:1 ratio and incubated at 37 °C for the indicated times. Reactions were stopped by adding 4× LDS sample buffer (Invitrogen), resolved by SDS–PAGE on 4–12% gradient gels run in MES buffer (Invitrogen) and visualized by silver staining.

Assembly of Ub chains for *in vitro* DUB assays and pull-down assays. Lys11-linked diUb variants were assembled using UBE2S¹⁹. To generate Ub moiety-specific mutations, such as Lys11 diUb with a proximal K33A or K33E mutation, chains were assembled using Ub (K11R, K63R) as the distal, and Ub (K63R, ΔLRGG) variants as the proximal moiety. Lys11 diUb molecules carrying no further mutation are referred to as wild-type* (WT*), and diUb substrates carrying additional proximal mutations are called 'K33A*' and 'K33E*', respectively.

In order to generate specific branched triUb molecules containing a Lys11 linkage, the UBE2S assembly system was combined with UBE2N/UBE2V1 or UBE2R1 for Lys11/63 or Lys11/48 triUb, respectively²⁸. Using Ub (ΔLRGG) with or without K63R mutation as the proximal Ub allowed the assembly of defined branches. Furthermore, distal Ub moieties contained mutations to prevent chain elongation, that is, Ub (K11R, K63R) and Ub (K11R, K48R, K63R) were used for Lys11/63 and Lys11/48-branched triUb, respectively.

FRET-based DUB kinetics. Recently developed FRET-based diUb substrates were used to determine Michaelis–Menten kinetics for Cezanne variants²⁰. Two Ub moieties linked via a native isopeptide linkage feature a donor (5-carboxyrhodamine110; Rho) or an acceptor (5-carboxytetramethylrhodamine; TAMRA) fluorophore, respectively. Lys11- and Lys48-linked diUb substrates of this type were used, while Lys63 FRET diUb was purchased from Boston Biochem (cat. no. UF-330). The change in fluorescence intensity (FI) of the donor fluorophore upon diUb hydrolysis was measured on a PheraStar plate reader (BMG Labtech), equipped with an FI optic module with λ_{ex} = 485 nm and λ_{em} = 520 nm (for Lys11 and Lys48 FRET diUb), or λ_{ex} = 540 nm and λ_{em} = 590 nm (for Lys63 FRET diUb). Reactions were performed in black, round-bottomed, non-binding 384-well plates (Corning) at 25 °C in a total volume of 15 μl.

Non-fluorescent Lys11, Lys63 and Lys48 diUb molecules were serially diluted in FI buffer (20 mM Tris (pH 7.5), 100 mM NaCl, 2 mM DTT, 0.1 mg/ml BSA) and the respective FRET substrate was spiked in at a fixed concentration of 1 μM (2× final concentration). In each well, 7.5 μl substrate was mixed with 7.5 μl enzyme at 2× indicated concentrations in FI buffer. All measurements for one experiment were performed in parallel in triplicate for each substrate concentration, and the change in FI was recorded over a period of 30–45 min at 15-s intervals. The maximal FI change was determined by using 25–50 nM Cez WT (Lys11 cleavage), 5 nM USP21 (Lys63 cleavage) or 0.5 μM OTUB1 (Lys48 cleavage) as positive controls. The observed FI values were plotted against time, and initial velocities of diUb cleavage were calculated. These initial rates at a fixed Cezanne concentration were plotted against diUb substrate concentration, and Michaelis–Menten parameters were determined using Prism 6 (GraphPad Software, Inc.). In the case of Lys48 diUb cleavage by Cez E157K, where the determined K_M value exceeded the highest tested diUb concentration, catalytic efficiencies were also calculated from a linear fit of the lower substrate concentration range (Extended Data Fig. 1d).

Modification of Cezanne with Ub-based suicide probes. Cezanne variants were diluted in DUB dilution buffer (25 mM Tris (pH 7.5), 150 mM NaCl, 10 mM DTT) to 1 l μM stocks, and mixed with 44 μM diUb ABPs²¹ in a 1:1 ratio. Reactions were incubated at 37 °C for 10–60 min, stopped with 4× LDS sample buffer (Invitrogen) and analysed on a Coomassie-stained SDS–PAGE gel.

Further Ub-based suicide probes used in this study include the Ub-haloalkyl probe Ub-C2Br²⁹ and Ub propargylamide (Ub-PA)³⁰, which were used for the generation of Cez–Ub for crystallization (Ub-C2Br) as well as A20–Ub crystallization, fluorescence polarization-based Ub-binding assays, NMR and HDX-MS (Ub-PA), respectively.

Crystallization. Crystallization screening was carried out in 96-well plates in a sitting-drop vapour diffusion setup using nano-litre robotics (typical drop size was 100+100 nl).

The first crystallized Cez apo construct (residues 88–438) contained an N-terminal flexible extension of 41 residues that was removed in subsequent crystallization attempts for Cez apo and complexes. Native (9.5 mg/ml) and selenomethionine (SeMet)-substituted (7.0 mg/ml) Cez apo (residues 88–438) crystals grew at 18 °C in 8% (w/v) PEG 8K, 0.1 M lithium chloride and 50 mM magnesium sulphate. Cez apo (residues 129–438, 8.0 mg/ml) crystallized in 0.1 M Bis-Tris pH 6.1 and 0.2 M magnesium formate at 18 °C.

Covalent complexes of Cezanne and A20 were generated with Ub-derived suicide probes (see above) and purified by anion-exchange chromatography and gel filtration. To obtain crystals of Cez–Ub, the long, unstructured V-loop (residues 267–291) was replaced by the corresponding sequence in TRABID (Gln-Pro-Gly; QPG). Cez (residues 129–438, QPG) was reacted with Ub-C2Br to form Cez–Ub, which was set up at a concentration of 21.7 mg/ml. Initial crystals grew from 0.1 M sodium acetate pH 4.6 and 8% (w/v) PEG 4K at 18 °C, and were used for streak seeding to obtain diffraction-quality crystals that grew from 0.1 M sodium acetate pH 4.8 and 6% (w/v) PEG 6K. For Cez–Lys11 diUb crystallization, Cez (residues 129–438) was reacted with Lys11-linked diUb ABP. The complex (6.7 mg/ml) crystallized at 18 °C in 0.1 M phosphate citrate pH 4.2, 20% (w/v) PEG 8K and 0.2 M sodium chloride. A20–Ub was generated from A20 (residues 1–366) and Ub-PA. Crystals were set up at a concentration of 8.0 mg/ml and grew from 0.1 M MES/imidazole pH 6.5, 7% (w/v) PEG 8K and 20% ethylene glycol at 14 °C.

Prior to synchrotron data collection, crystals were vitrified in liquid nitrogen after brief soaking in mother liquor containing 27.5% glycerol (Cez apo, residues 88–438), 25% (v/v) PEG400 (Cez apo, residues 129–438), 30% (v/v) PEG400 (Cez–Ub), 28% glycerol (Cez–Lys11 diUb), or mother liquor alone (A20–Ub).

Data collection, structure determination and refinement. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF), beam lines ID23-1 (Cez apo; wavelength: 0.97933 Å) and ID29 (A20–Ub; 0.96863 Å), the Diamond Light Source, beam lines I02 (Cez–Ub; 0.97950 Å) and I03 (Cez–Lys11 diUb; 0.97626 Å), and the High Energy Accelerator Research Organization (KEK), beam line PF BL-1A (Cez SeMet; 0.9786 Å).

Diffraction images were indexed and integrated using iMOSFLM³¹ or XDS³², and scaled using SCALA³³ or its successor program AIMLESS³⁴.

The structure of Cez apo (residues 88–438) was solved by SAD phasing using diffraction data collected from a SeMet-substituted crystal. The automated structure solution pipeline SHARP and autoSHARP were used for SAD phasing^{35,36}, followed by iterative manual building using Coot³⁷ and refinement using PHENIX³⁸. Electron density was not visible for the first 41 residues, which were removed from the construct for subsequent crystallization attempts.

Phases for subsequent Cez structures and for A20–Ub were obtained by molecular replacement in PHASER³⁹, using Cez apo (88–438), Cez apo (129–438) or A20 apo (PDB 2VFJ²²), and Ub (PDB 1UBQ⁴⁰), as search models where appropriate. Models were built in Coot³⁷ and refined in PHENIX³⁸ in iterative rounds, using simulated annealing and TLS restraints where appropriate. Final

Ramachandran statistics (favoured/allowed/outliers) were 98.5%/1.5%/0% (Cez apo), 99.1%/0.9%/0% (Cez–Ub), 98.6%/1.4%/0% (Cez–Lys11 diUb), and 97.1%/2.9%/0% (A20–Ub), respectively. Final data collection and refinement statistics are summarized in Extended Data Table 1.

Fluorescence polarization assays. Fluorescence polarization measurements were used to monitor the interaction between DUBs and FLAsH-tagged Ub, as well as the cleavage of a monoubiquitinated TAMRA-labelled Lys–Gly (KG) dipeptide (Ub–KG)*²⁴ or Lys11-linked diUb–FLAsH. Measurements were performed using a PheraStar plate reader (BMG Labtech), equipped with an optic module for FLAsH ($\lambda_{\text{ex}} = 485 \text{ nm}$, $\lambda_{\text{em}} = 520 \text{ nm}$) or TAMRA ($\lambda_{\text{ex}} = 540 \text{ nm}$ and $\lambda_{\text{em}} = 590 \text{ nm}$) detection. Fluorescence polarization values given in millipolarization (mP) were determined by taking the following polarization values into account: FLAsH–Ub (195 mP), TAMRA–KG (50 mP) and Lys11 diUb–FLAsH (315 mP). Assays were performed in triplicate in black, round bottom, non-binding surface 384-well plates (Corning) at 25 °C in 20 μL .

For fluorescence polarization binding studies, FLAsH–Ub was diluted to 300 nM in fluorescence polarization buffer (20 mM Tris (pH 7.5), 100 mM NaCl, 2 mM β -mercaptoethanol, 0.1 mg/ml BSA). Serial dilutions in fluorescence polarization buffer were prepared of Cezanne variants, and 10 μL of this was mixed with 10 μL FLAsH–Ub in a 384-well plate. The plate was incubated in the dark for 10–15 min before the measurement. Fluorescence polarization values were fitted according to a ‘one-site – total’ binding model using Prism 6 (GraphPad Software, Inc.).

Fluorescence polarization cleavage assays were started by mixing 10 μL per well enzyme at 2 \times indicated concentration in fluorescence polarization buffer to 10 μL of predisposed fluorescence polarization substrate (Ub–KG* or Lys11-linked diUb–FLAsH) at 300 nM. For the TAMRA-based substrate, TAMRA–KG was included as positive control. Hydrolysis reactions were followed for 1 h in 60–90 s intervals. Fluorescence polarization values of TAMRA-measurements were fitted according to a ‘one phase decay’ model using Prism 6 (GraphPad Software, Inc.).

NMR Ub-binding study with wild-type Cez apo and Cez–Ub. Cez WT (residues 129–438), a purified covalent Cez–Ub complex (assembled with Cez WT and Ub–PA), and ¹⁵N-labelled Ub WT were dialysed against NMR buffer (18 mM Na₂HPO₄, 7 mM NaH₂PO₄·xH₂O, 150 mM NaCl, 5 mM DTT (pH 7.2)). Samples of 50 μM ¹⁵N-labelled Ub alone or in the presence of 130 μM Cez WT or Cez–Ub were prepared in 350 μL NMR buffer containing 20 μL D₂O, and were transferred into Shigemi NMR microtubes.

¹H–¹⁵N BEST-TROSY (band-selective excitation short-transient transverse relaxation-optimized spectroscopy) spectra were recorded at 298 K on a Bruker AvanceII+ 700 MHz spectrometer with a TCI triple resonance probe. Data were processed in TopSpin 3.0 (Bruker Inc.) and analysed in Sparky (UCSF).

Thermal shift assay. Protein melting curves were recorded on a Corbett RG-6000 real-time PCR cycler. Samples contained 3 μM Cez (residues 129–438) WT or C194A, 3 \times SYPRO Orange, and 0–400 μM Ub. Data were recorded in triplicate and in two independent experiments. Melting temperatures (T_m) in the presence of Ub were referenced to 45.9 \pm 0.2 °C (WT) and 45.0 \pm 0.2 °C (C194A).

Pull-down assay. His-tagged Cezanne constructs were used for *in vitro* pull-downs of purified Ub and diUb variants. For each reaction, 5 μL TALON Superflow resin (GE Healthcare) was equilibrated with PD buffer (20 mM Tris (pH 7.5), 100 mM NaCl, 2 mM β -mercaptoethanol), mixed with 20 μg His-tagged bait in 400 μL PD buffer plus 0.1 mg/ml BSA, incubated for 20 min at 4 °C under constant agitation, and subsequently washed three times. Ub and diUb prey proteins were diluted in PD buffer plus BSA to 1.2 $\mu\text{g}/\text{mL}$ and 2.4 $\mu\text{g}/\text{mL}$, respectively, and 400 μL was added to the immobilized bait. After incubation for 1 h at 4 °C, the resin was washed three times with PD buffer before the addition of 50 μL of 4 \times LDS sample buffer (Invitrogen). Samples were boiled for 1 min, and 20 μL per sample was analysed by SDS–PAGE and silver staining.

Hydrogen–deuterium exchange mass spectrometry (HDX–MS). Deuterium exchange reactions of Cezanne were initiated by diluting the protein in D₂O (99.8% D₂O ACROS, Sigma) in 10 mM Tris pH 7.5 buffer to give a final D₂O percentage of 95.3%. For all experiments, deuterium labelling was carried out at 23 °C (unless otherwise stated) at four time points (3 s on ice (0.3 s), 3 s, 30 s and 300 s). The labelling reaction was quenched by the addition of chilled 2.4% (v/v) formic acid in 2 M guanidinium hydrochloride, and immediately frozen in liquid nitrogen. Samples were stored at –80 °C before analysis.

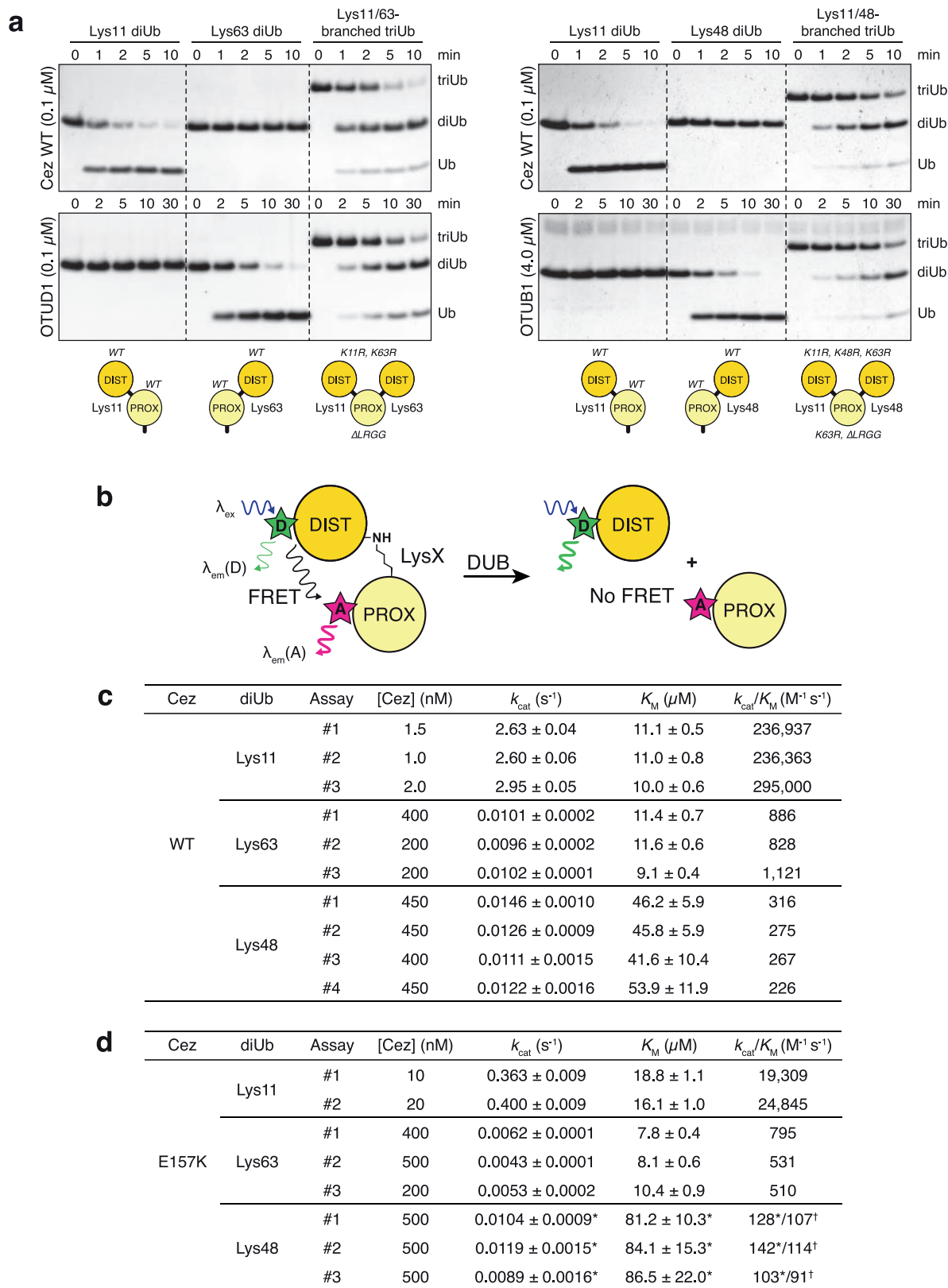
The quenched protein samples were rapidly thawed and subjected to proteolytic cleavage by pepsin followed by reversed phase HPLC separation. Briefly,

the protein was passed through an Enzymate BEH immobilized pepsin column, 2.1 \times 30 mm, 5 μm (Waters, UK) at 200 $\mu\text{L}/\text{min}$ for 2 min and the peptic peptides trapped and desalted on a 2.1 \times 5 mm C18 trap column (Acquity BEH C18 Vanguard pre-column, 1.7 μm , Waters). Trapped peptides were subsequently eluted over 12 min using a 5–36% gradient of acetonitrile in 0.1% (v/v) formic acid at 40 $\mu\text{L}/\text{min}$. Peptides were separated on a reverse phase column (Acquity UPLC BEH C18 column 1.7 μm , 100 mm \times 1 mm (Waters). Peptides were detected on a SYNAPT G2–Si HDMS mass spectrometer (Waters) acquiring over an m/z of 300–2,000, with the standard electrospray ionization (ESI) source and lock mass calibration using [Glu1]–fibrinopeptide B (50 fmol/ μL). The mass spectrometer was operated at a source temperature of 80 °C and a spray voltage of 2.6 kV. Spectra were collected in positive ion mode.

Peptide identification was performed by MS⁺ using an identical gradient of increasing acetonitrile in 0.1% (v/v) formic acid over 12 min. The resulting MS⁺ data were analysed using Protein Lynx Global Server software (Waters) with an MS tolerance of 5 ppm.

Mass analysis of the peptide centroids was performed using DynamX software (Waters). Only peptides with a score >6.4 were considered. The first round of analysis and identification were performed automatically by the DynamX software, however, all peptides (deuterated and non-deuterated) were manually verified at every time point for the correct charge state, presence of overlapping peptides, and correct retention time. Deuterium incorporation was not corrected for back-exchange and represents relative, rather than absolute changes in deuterium levels. Changes in H–D amide exchange in any peptide may be due to a single amide or a number of amides within that peptide. All time points of a data set (that is, data of related constructs) were prepared simultaneously and were acquired on the mass spectrometer on the same day.

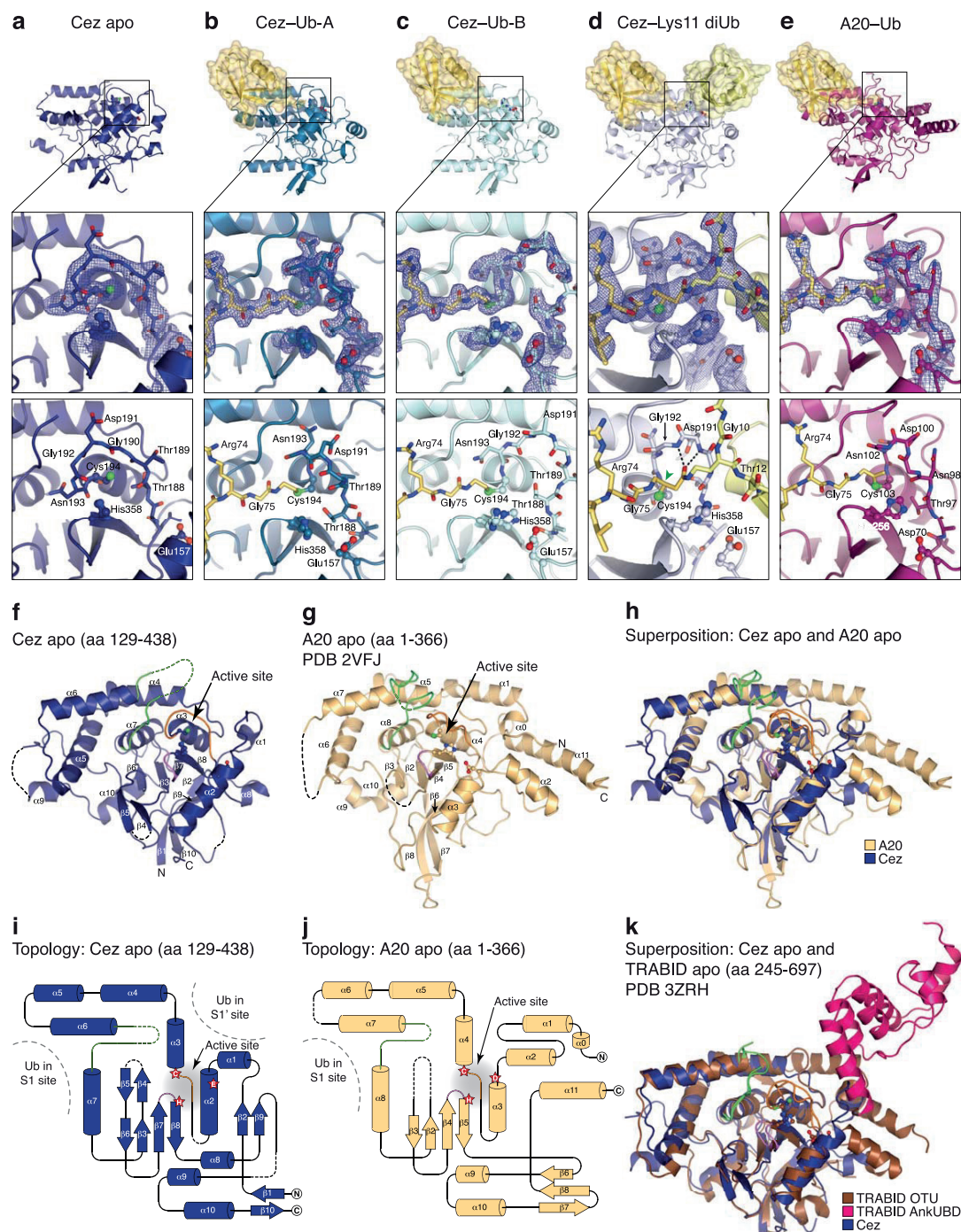
26. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
27. Hospenthal, M. K., Mevissen, T. E. T. & Komander, D. Deubiquitinase-based analysis of ubiquitin chain architecture using Ubiquitin Chain Restriction (UbiCRest). *Nat. Protocols* **10**, 349–361 (2015).
28. Faggiano, S., Alfano, C. & Pastore, A. The missing links to link ubiquitin: Methods for the enzymatic production of polyubiquitin chains. *Anal. Biochem.* **492**, 82–90 (2016).
29. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159 (2002).
30. Ekkebus, R. *et al.* On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **135**, 2867–2870 (2013).
31. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
32. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
33. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
34. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
35. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030 (2003).
36. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
38. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
39. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
40. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).
41. Pham, G. H. *et al.* Comparison of native and non-native ubiquitin oligomers reveals analogous structures and reactivities. *Protein Sci.* **25**, 456–471 (2016).
42. Messick, T. E. *et al.* Structural basis for ubiquitin recognition by the Otu1 ovarian tumor domain protein. *J. Biol. Chem.* **283**, 11038–11049 (2008).
43. Wiener, R. *et al.* E2 ubiquitin-conjugating enzymes regulate the deubiquitinating activity of OTUB1. *Nat. Struct. Mol. Biol.* **20**, 1033–1039 (2013).



Extended Data Figure 1 | See next page for caption.

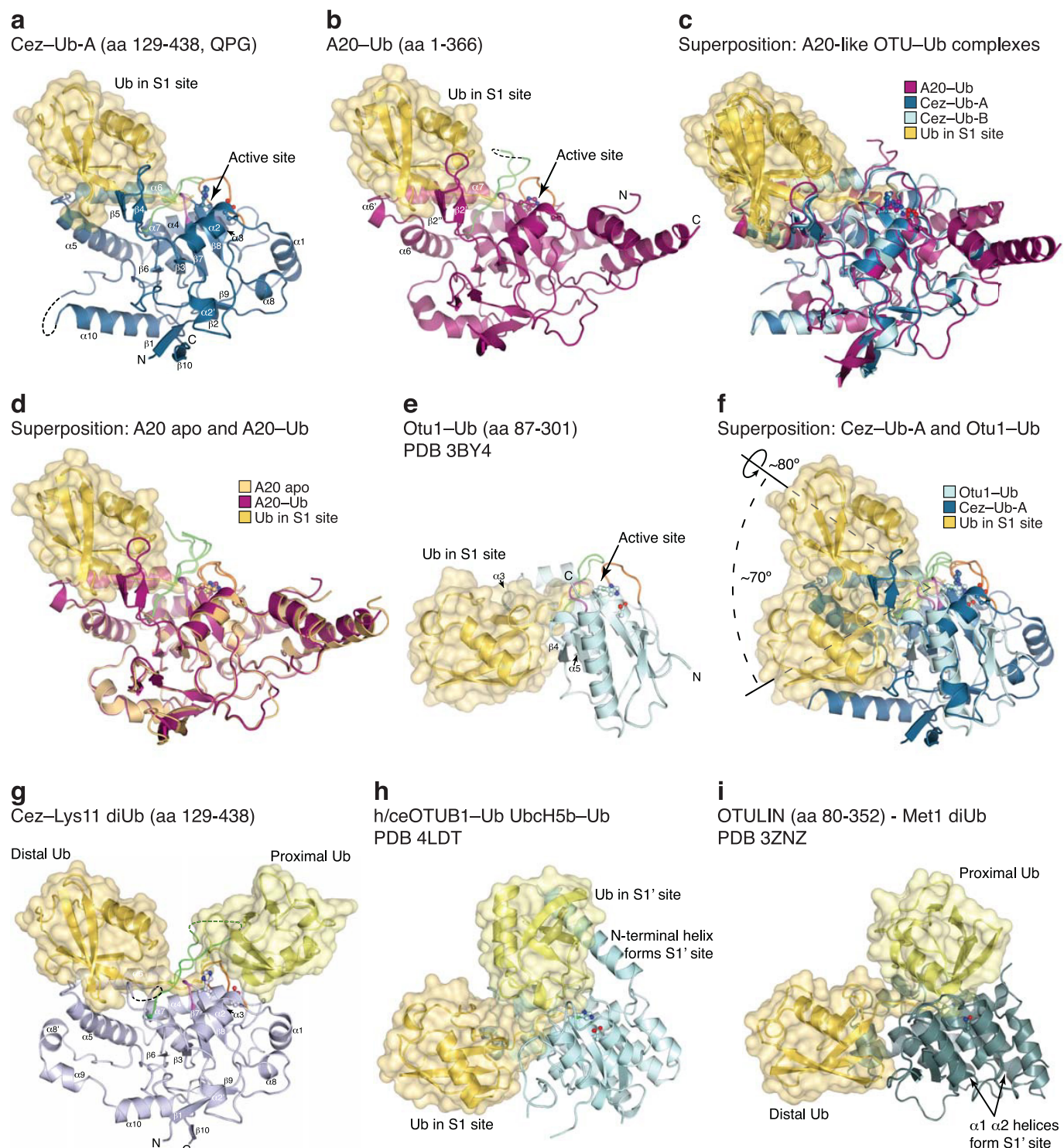
Extended Data Figure 1 | Analysis of branched triUb substrates and FRET-based diUb cleavage kinetics. **a**, Branched triUb molecules with different topologies were generated as shown in the schematic (bottom). Lys11 diUb, Lys63 diUb and branched Lys11/63 triUb (left panel) were treated with wild-type Cezanne (Cez WT; top) and OTUD1 (residues 287–481, bottom), a Lys63-specific enzyme². Both DUBs cleaved their preferred diUb substrate as well as one linkage of the branched triUb molecule. Lys11 diUb, Lys48 diUb and branched Lys11/48 triUb (right panel) were incubated with wild-type Cezanne (top) and OTUB1 (full-length, bottom), a Lys48-specific OTU DUB². Again, both enzymes showed similar activities towards their preferred linkage type in a diUb substrate and a branched triUb molecule. This shows that Cezanne can cleave Lys11 linkages in the context of Lys11/Lys63- and Lys11/Lys48-branched chains. For gel source images, see Supplementary Fig. 1. **b**, Schematic of FRET-based diUb cleavage assays to derive DUB kinetics. Distal and proximal Ub moieties were modified with a donor (D) and acceptor (A) fluorophore, respectively. Upon DUB treatment, the native

isopeptide bond was cleaved and the FRET signal was lost. The increase in donor intensity was measured to follow the reaction. **c**, Kinetic parameters for all independently performed experiments of Lys11, Lys63 and Lys48 diUb cleavage by wild-type Cezanne. Values are in good agreement with previously published parameters derived from gel-based studies⁴¹. **d**, Summary of kinetic parameters for Lys11-, Lys63- and Lys48-linked diUb cleavage by Cezanne E157K. The determined K_M values for Lys48 diUb lie above the highest tested substrate concentration, so kinetic parameters marked by an asterisk were calculated from experiments where substrate saturation could not be achieved owing to technical limitations. Catalytic efficiencies (k_{cat}/K_M) for this substrate were also derived from a linear fit of the lower concentration range (0–20 μM , linear part of the graph). These values are marked by a cross. The similarity of catalytic efficiencies calculated in two different ways indicate that the kinetic parameters marked by asterisks are good estimates. See Supplementary Fig. 2 for all corresponding graphs of initial rates.



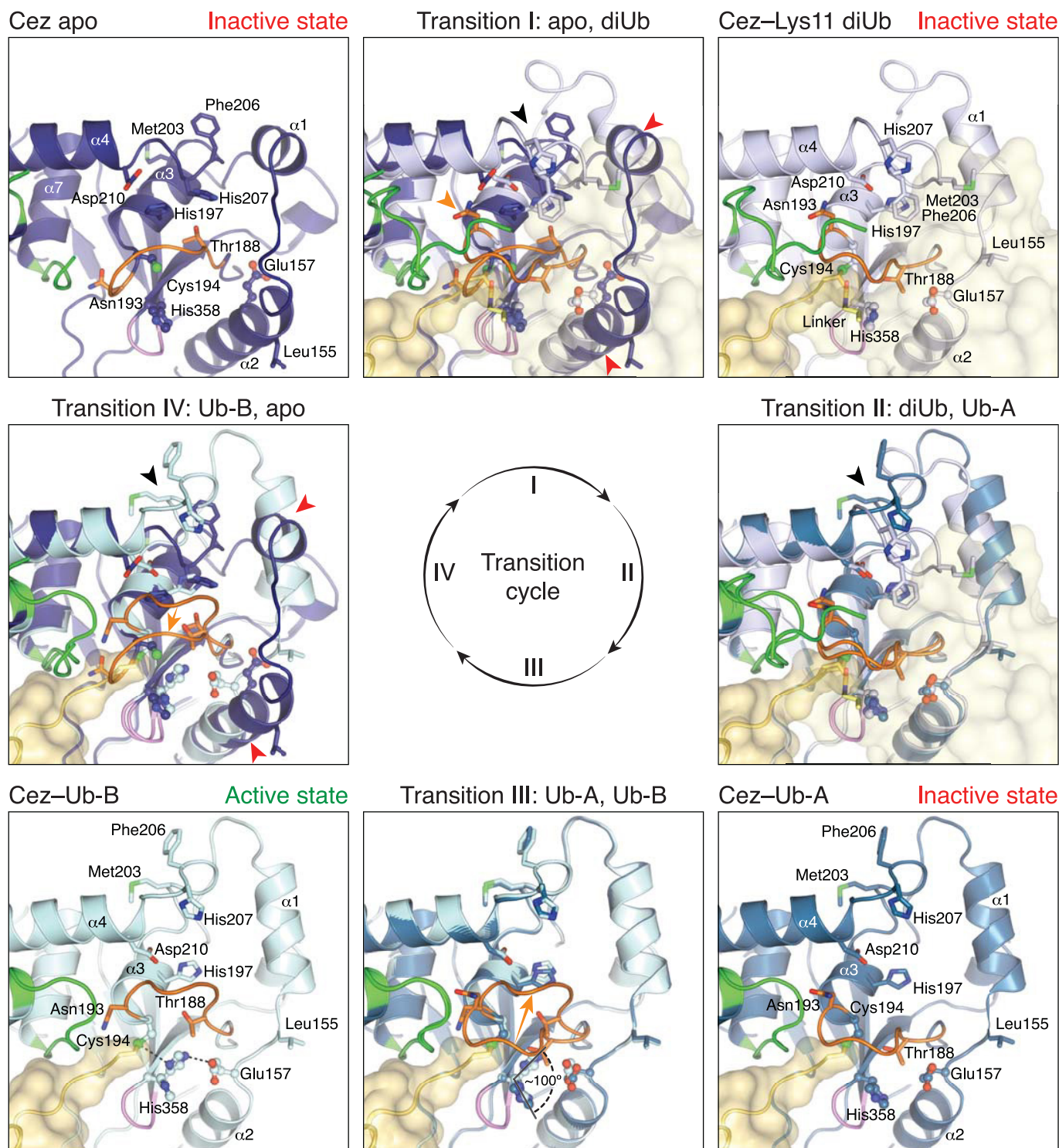
Extended Data Figure 2 | Crystal structures determined in this study and comparison of A20-like OTU apo structures. **a–e**, Active site regions Cez apo (**a**), Cez-Ub-A (**b**), Cez-Ub-B (**c**), Cez-Lys11 diUb (**d**), and A20-Ub (**e**). $2|F_o| - |F_c|$ electron density maps contoured at 1σ (blue) cover catalytic residues, the Cys-loop and chemical linkers in the complex structures. Hydrogen bonds between the oxyanion hole and the Lys11 diUb ABP linker carbonyl are indicated in **d**, and the sp^3 -hybridized carbon atom that is linked to the oxyanion in a native first tetrahedral intermediate is highlighted (green arrowhead). **f**, **g**, Cezanne OTU (as in Fig. 1d) and A20 OTU (PDB 2VFJ)²² apo structures with labelled secondary structure elements. Catalytic residues are shown in ball-and-

stick representation. Three loops surrounding the active site are coloured (Cys-loop, orange; V-loop, green; His-loop, purple). **h**, Superposition of **f** and **g** showing structural similarities and differences between Cezanne and A20. **i**, **j**, Topology diagrams of **f** and **g**. The catalytic centre is indicated (red stars) and Ub-binding sites are highlighted. A20 contains two additional N-terminal and one additional C-terminal helices compared to Cezanne. The $\beta 1$ – $\beta 10$ sheet in Cezanne corresponds to the A20 $\beta 7$ – $\beta 8$ sheet. This explains why sequence-based alignments are challenging. **k**, Superposition of Cez apo (**f**) and TRABID AnkUBD (pink) and OTU (brown) domains (residues 245–697, PDB 3ZRH¹⁰).



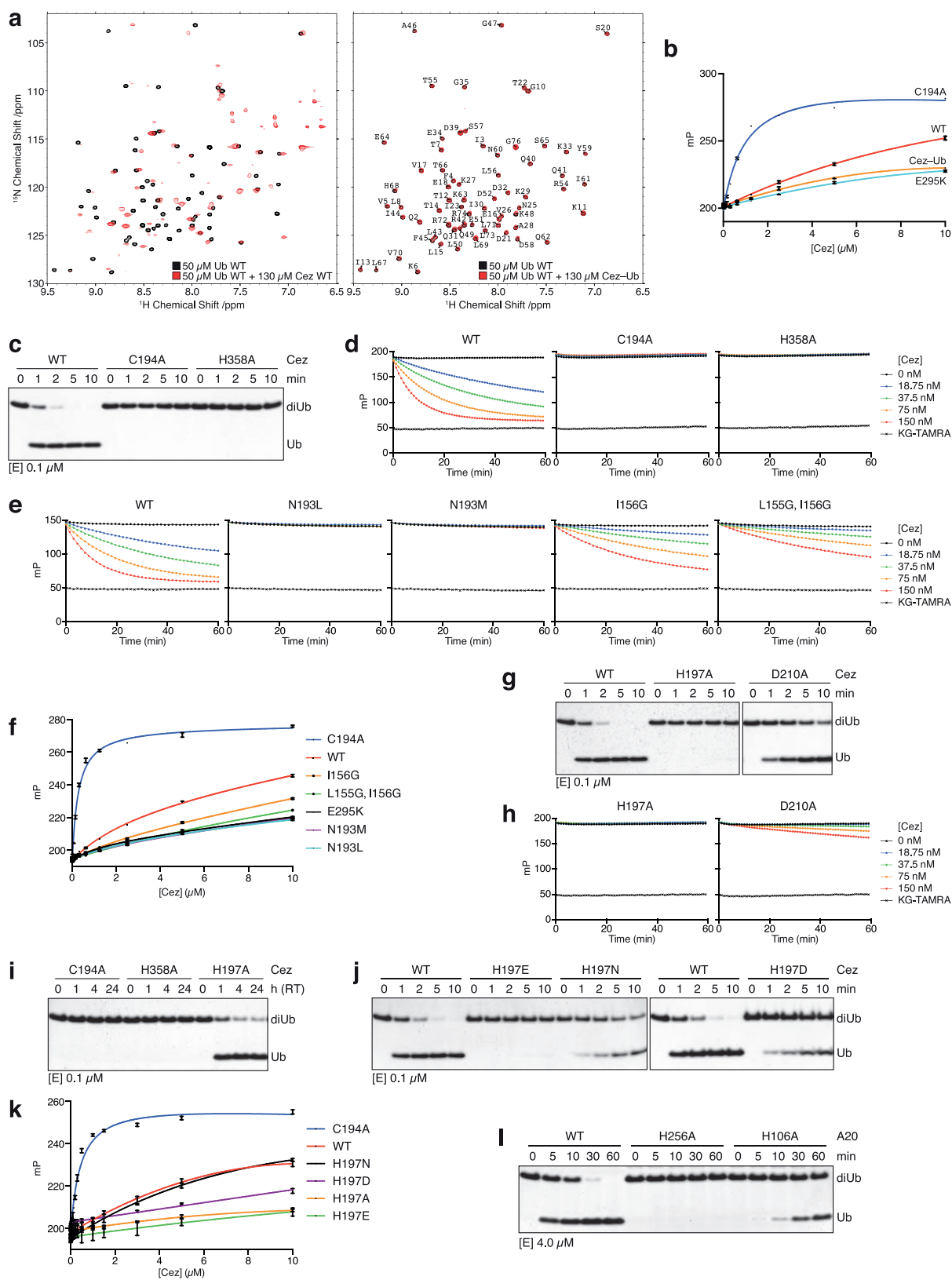
Extended Data Figure 3 | Comparison of Ub and diUb complexes within the OTU family. Ub moieties are shown in cartoon representation under transparent surfaces in shades of yellow. Secondary structure elements involved in Ub binding are labelled, and active site loops are coloured as in Extended Data Fig. 2f. **a**, Cez–Ub–A complex as in Fig. 1d. **b**, A20–Ub complex as in Fig. 1g. **c**, Superposition of Ub complexes reveals a conserved S1 Ub-binding mode in A20-like OTU DUBs. **d**, Superposition of A20 apo (Extended Data Fig. 2g) and A20–Ub (**b**). No large conformational changes occur upon Ub binding. However, two unstructured loops in A20 apo are stabilized by Ub, forming helix $\alpha 6'$ and the $\beta 2'-\beta 2''$ sheet (compare with Extended Data Fig. 2j). **e**, The structure of the yeast Otu1–Ub complex (PDB 3BY4⁴²) is representative of the

OTUD subfamily of OTU DUBs. The Ub moiety in the S1 site is mainly bound by the short helix $\alpha 3$. **f**, The superposition of Cez–Ub (**a**) and Otu1 (**e**) reveals substantially different S1 sites between A20-like and OTUD subfamilies. Rotations around the roll axis of Ub ($\sim 80^\circ$) and the active site ($\sim 70^\circ$) would be required to align both Ub moieties. **g–i**, Structures of OTU domains in identical orientation bound to their respective diUb substrate. The binding modes of proximal and distal Ub differ markedly between the here determined Cez–Lys11 diUb complex (**g**, as shown in Fig. 1d), the h/ceOTUB1–Ub Ubch5b–Ub structure (PDB 4LDT⁴³; Ubch5b molecule is not shown), which resembles an OTUB1–Lys48 diUb complex (**h**), and OTULIN bound to Met1-linked diUb (PDB 3ZNZ³) (**i**).



Extended Data Figure 4 | Conformational changes in the catalytic centre. Cezanne structures (as in Fig. 2a) are shown in the corners, and transitions I–IV are overlays of neighbouring structures. Side chains of catalytic residues and other selected residues are highlighted. Loops are coloured as in Extended Data Fig. 2f. Cez apo shows a catalytically incompetent state. His358 and Glu157 are in flipped-out conformations. Transition I features structural rearrangements of the Cys-loop (orange arrowhead), helices $\alpha 1$ and $\alpha 2$ (red arrowheads) and the S1'-loop (black arrowhead). Cez–Lys11 diUb also features an inactive state; His358

remains flipped-out, which is caused by the Cys-loop residue Thr188 that is pushed into the active site by the proximal Ub. In transition II, another S1'-loop movement relocates S1' site residues (black arrowhead). A similar inactive state is present in Cez–Ub-A, and Thr188 still resides in the active site. The absence of the proximal Ub allows the Cys-loop and Thr188 to move in transition III (orange arrow), allowing a $\sim 100^\circ$ rotation of His358. Hence, Cez–Ub-B contains an aligned catalytic centre. Hydrogen bonds are indicated. In transition IV, large conformational changes in various parts of the OTU domain regenerate the autoinhibited apoenzyme.

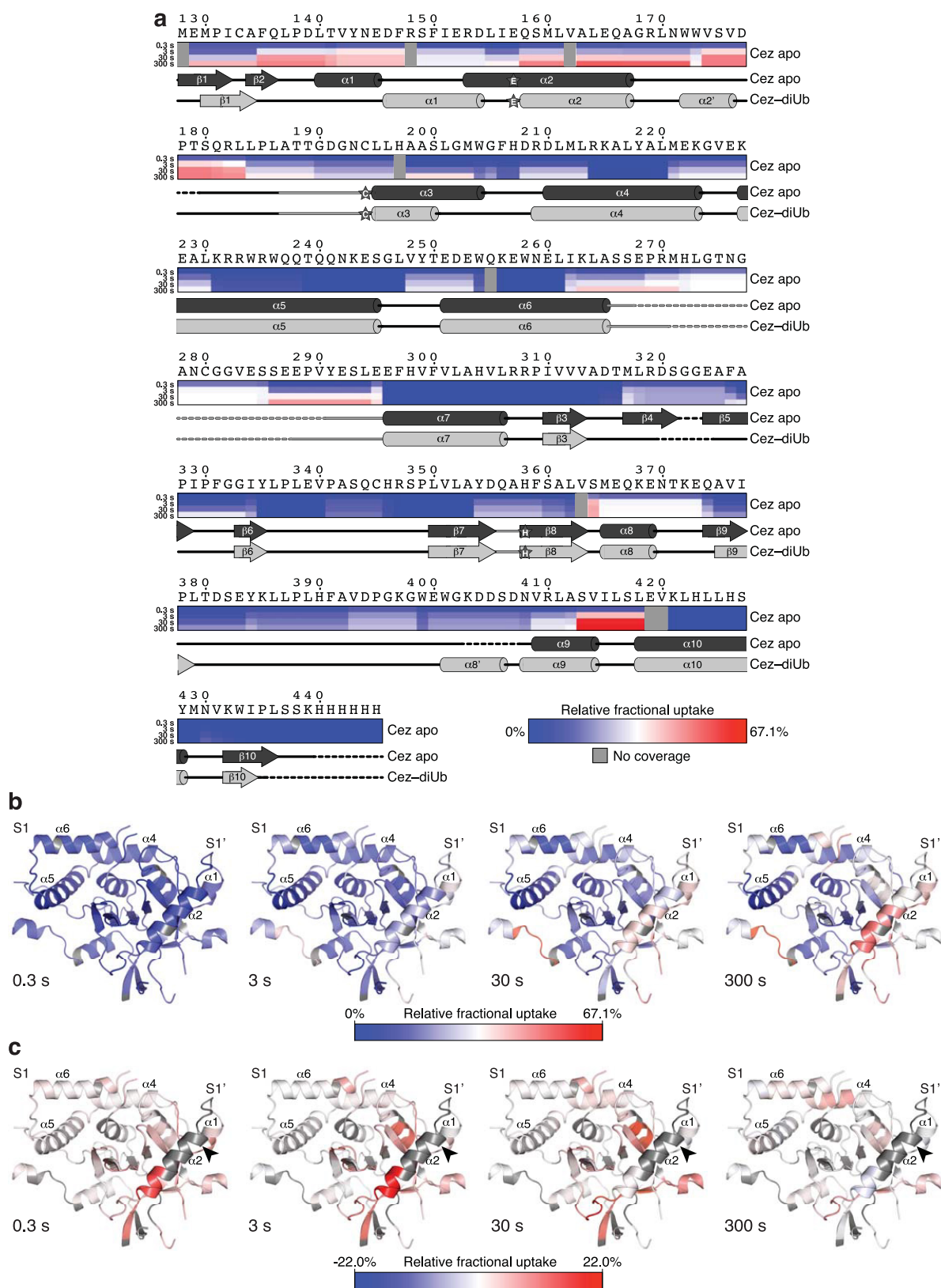


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Ub binding to Cezanne and mutational analysis of residues involved in catalysis and conformational dynamics.

a, NMR analysis of Ub binding to wild-type Cezanne and the covalent Cez–Ub complex. ^1H - ^{15}N BEST-TROSY spectra of $50\ \mu\text{M}$ ^{15}N -labelled Ub alone (black) and in the presence of $130\ \mu\text{M}$ unlabelled wild-type Cezanne (red, left) or unlabelled Cez–Ub (red, right). Strong chemical shift perturbations upon addition of wild-type Cezanne indicate binding to Ub. In contrast, no chemical shifts were detected with Cez–Ub, suggesting that all changes with wild-type Cezanne can be attributed to the S1 site (this site is occupied by unlabelled Ub in Cez–Ub). More importantly, this also indicates that a functional S1' site is not present in the Cez–Ub complex. **b**, Fluorescence polarization experiment assessing the binding of FLAsH-tagged Ub to catalytically inactive Cezanne (C194A), wild-type Cezanne, an S1 site mutant (E295K, see below) and the Cez–Ub complex. **c**, Lys11 diUb cleavage assays of catalytic Cys194 and His358 mutants. **d**, **e**, Ub-KG* cleavage by catalytic Cys194 and His358 mutants (**d**), as well as Asn193 and helix $\alpha 2$ mutants that modulate the overall dynamics of Cezanne (**e**). This assay follows fluorescent dye release in the reaction; the fact that Cezanne H358A is inactive indicates an important role in the deprotonation of the catalytic Cys at the start of the reaction (that is, the catalytic centre transiently adopts an active state) and/or a role in resolving the first tetrahedral intermediate. If His358 was not required for either,

we would expect a single turnover of the reaction, which would stop at the thioester intermediate. The release of KG-TAMRA would still occur, but this was not detected even at an enzyme concentration of $150\ \text{nM}$ (the substrate concentration in all assays was $150\ \text{nM}$). The fluorescence polarization signal also did not increase, suggesting that no covalent first tetrahedral intermediate was formed due to impaired dye release. Hence, the data suggest a role for His358 at least in the initial Cys deprotonation in addition to the last reaction step. **f**, Fluorescence polarization binding assay of Asn193 and helix $\alpha 2$ mutants compared to constructs used in **b**. **g**, **h**, Hydrolysis of Lys11-linked diUb (**g**) and Ub-KG* (**h**) by Cezanne H197A and D210A. **i**, DUB assay with Cezanne variants (extended incubation at room temperature, RT). **j**, Lys11 diUb cleavage assay with His197 variants. **k**, Fluorescence polarization binding assay as in **b** testing His197 variants. **l**, Mutation of corresponding residues in A20 (A20 His256 corresponds to Cezanne His358, and A20 His106 to Cezanne His197) have similar effects on Lys48 diUb hydrolysis. All DUB assays are representative of at least two independent experiments for every construct. Ub-KG* cleavage experiments and fluorescence polarization binding assays were replicated at least twice for each variant with consistent results. Fluorescence polarization measurements were performed in triplicate. Error bars represent s.d. from the mean. mP, millipolarization unit. For gel source images, see Supplementary Fig. 1.

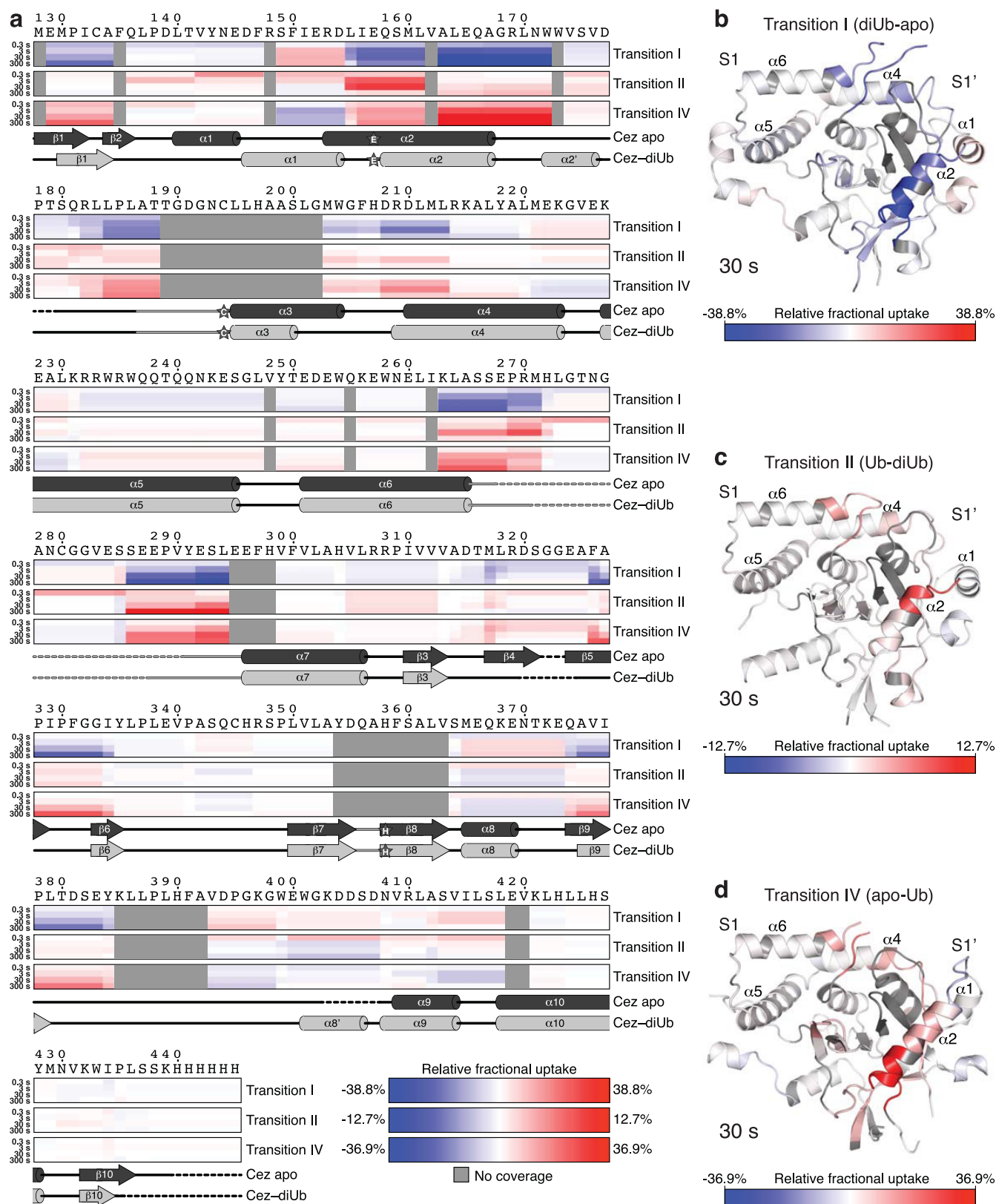


Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | HDX-MS analysis of the Cez apo state.

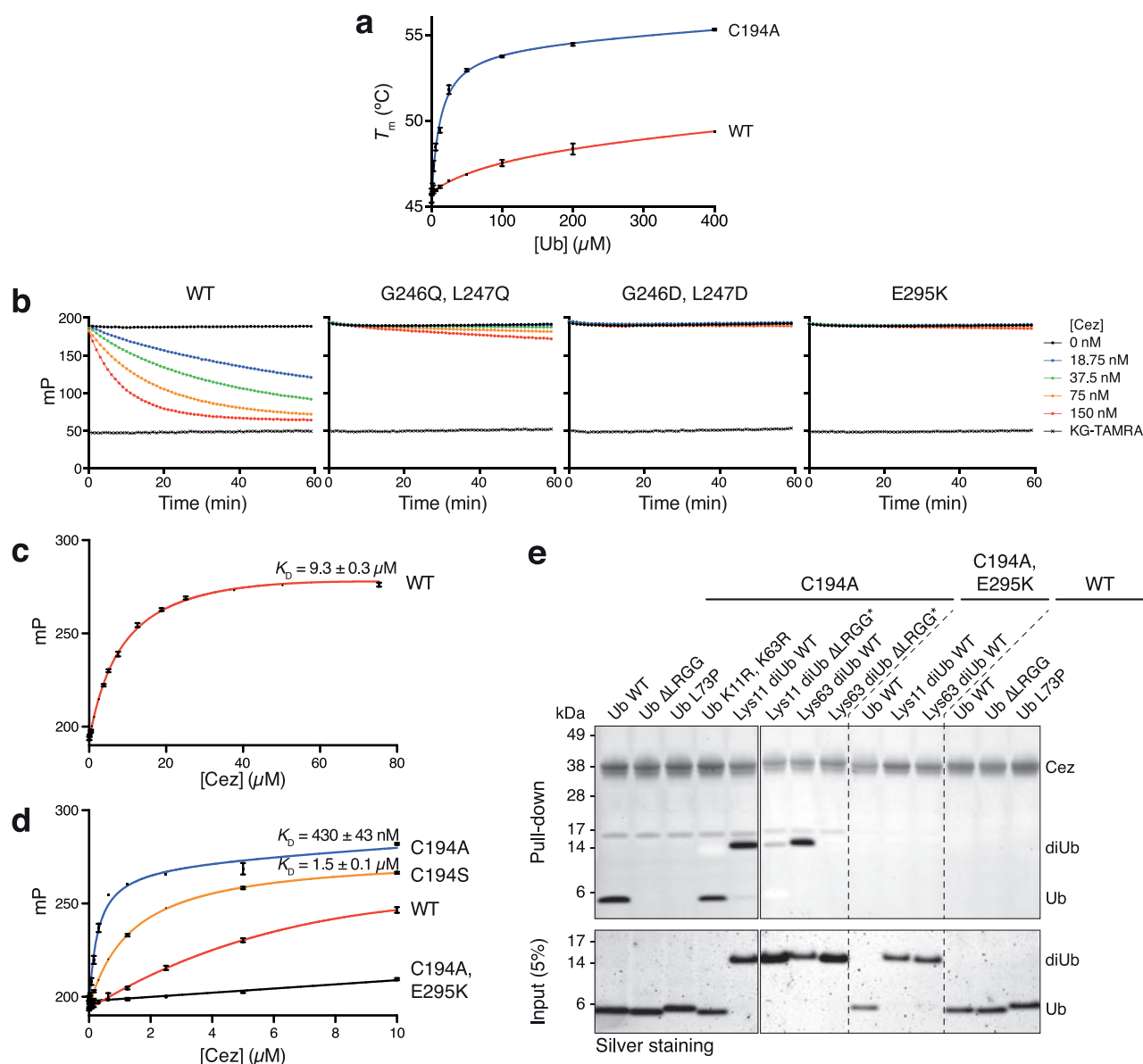
a, HDX-MS experiment showing the conformational dynamics of wild-type Cezanne. The relative fractional deuterium uptake is shown for four time points (0.3–300 s). Protein sequence and secondary structure elements of Cez apo (dark grey) and Cez–diUb (light grey) are aligned. Residues of the catalytic centre are indicated by stars. **b**, Cez apo structure coloured based upon the relative fractional deuterium uptake of wild-type Cezanne at 0.3 s, 3 s, 30 s and 300 s. The region spanning helices $\alpha 1$ and $\alpha 2$ shows a particularly high deuterium uptake, suggesting conformational flexibility in this region in solution. **c**, The H–D exchange of the $\alpha 2$ -helix destabilizing mutant Cezanne L155G/I156G compared to wild-type Cezanne. Cez apo structure coloured based upon the difference

in deuterium uptake (L155G/I156G-WT) at 0.3 s, 3 s, 30 s and 300 s (heat maps are shown in Supplementary Fig. 3). The data suggest that helix $\alpha 2$ is destabilized, as regions structurally adjacent to the mutation site (black arrowhead) show increased deuterium uptake as compared to wild-type Cezanne. Peptides containing the mutations could not be analysed owing to the different sequences, and are therefore coloured grey. Notably, most differences are stronger at shorter time points, indicating increased dynamics within this time frame (0.3–30 s). At the last time point (300 s), differences are not as pronounced, suggesting that wild-type Cezanne undergoes similar structural rearrangements at a slower speed. Importantly, the data also confirm that overall folding of the mutant was not affected by the two Gly residues introduced into helix $\alpha 2$.



Extended Data Figure 7 | HDX-MS analysis of transitions I, II and IV. **a**, HDX-MS experiments were performed with Cez apo, Cez-Lys11 diUb and Cez-Ub. Heat maps show differences in deuterium uptake between two states in each case: transition I (diUb-apo), transition II (Ub-diUb) and transition IV (apo-Ub). Hence, Cezanne regions that are stabilized or more protected upon Lys11 diUb binding (transition I), or more flexible or exposed upon the stepwise release of the proximal Ub (transition II) and the distal Ub (transition IV) are highlighted. The S1 site predominantly

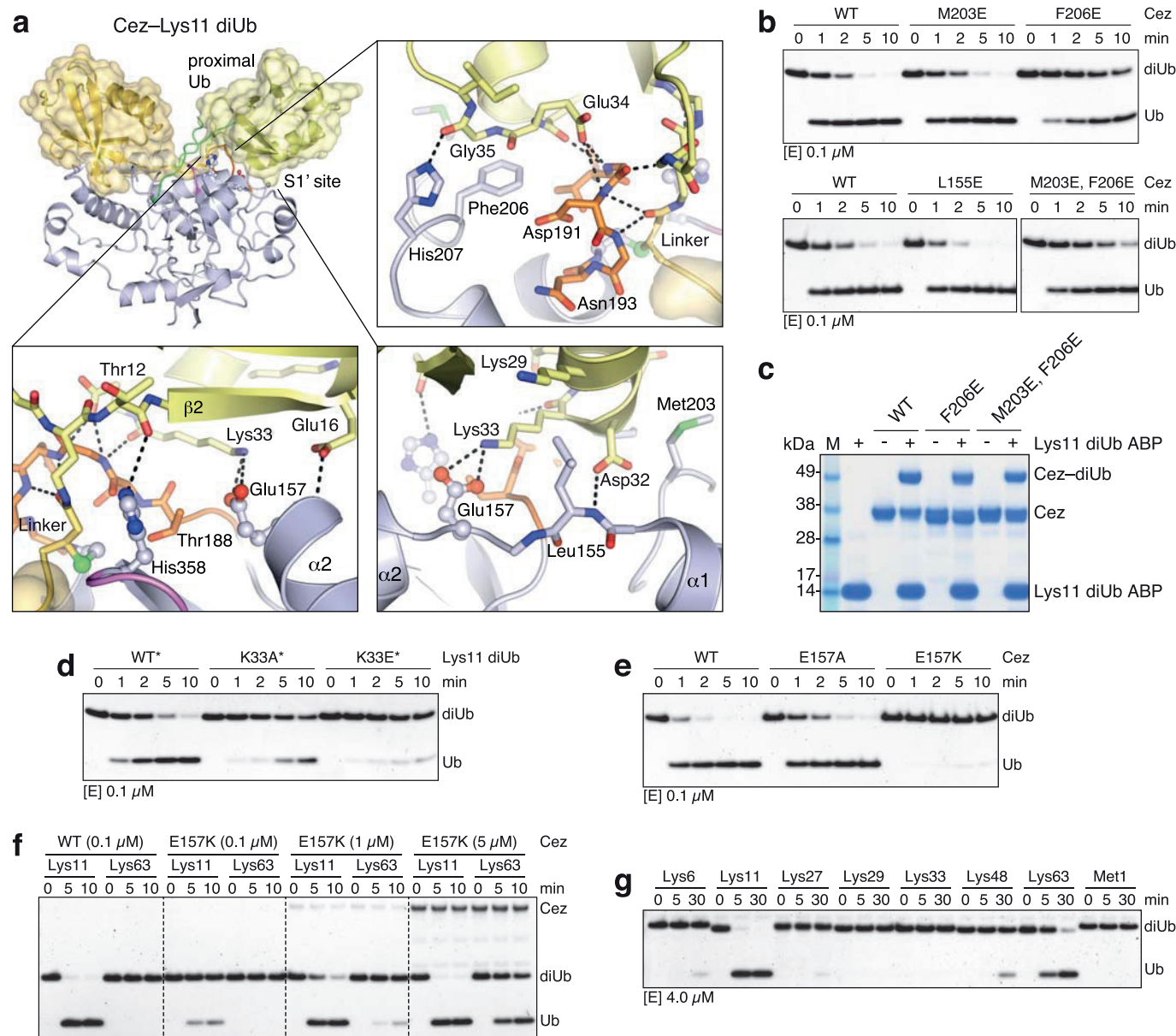
consists of helices $\alpha 5$ and $\alpha 6$ (that is, helical content with very low deuterium uptake in any state), and is not as easily detected as the S1' site, which features various loops and the dynamic helix $\alpha 2$. Cezanne sequence and secondary structure schematics are shown as in Extended Data Fig. 6a. **b**, Cez-Lys11 diUb structure (shown without Lys11 diUb) coloured based upon transition I deuterium uptake at 30 s. **c**, Transition II deuterium uptake at 30 s plotted onto Cez-Ub-B (shown without Ub). **d**, Cez apo coloured based upon transition IV deuterium uptake at 30 s.



Extended Data Figure 8 | Mutational analysis of the S1 Ub-binding site.

a, Thermal shift assay of wild-type and C194A Cezanne. In the presence of Ub, the melting temperature (T_m) of Cezanne increases. Data were recorded in triplicate and in two independent experiments. **b**, Ub-KG* hydrolysis by S1 site mutants. **c**, **d**, Fluorescence polarization-based affinity measurement using N-terminally FLAsH-tagged Ub. Dissociation constants (K_D) for wild-type (**c**), C194A and C194S Cezanne (**d**) are shown. Data are representative of at least two independent experiments per construct. **e**, Pull-down assay with His-tagged Cezanne constructs (catalytically inactive C194A, S1 site mutant C194A/E295K or

wild-type) and different Ub and diUb variants. MonoUb requires an intact C terminus to bind to Cezanne C194A. To prevent unspecific binding of differently linked diUb molecules with their proximal Ub to the S1 site, the C terminus was removed (Δ LRGG). Variants marked by an asterisk were assembled using K11R, S20C and K63R mutations in the distal Ub, as well as K63R (only for Lys11 diUb) and Δ LRGG in the proximal Ub moiety. Pull-down and input samples were analysed by SDS-PAGE and silver staining. The pull-down assay was performed in two independent experiments. For gel source images, see Supplementary Fig. 1.



Extended Data Figure 9 | Biochemical analysis of S1' site mutations.

a, The interface between Cezanne and the proximal Ub in the Cez–Lys11 diUb complex. An unusual surface of Ub comprising Glu16, Asp32, Lys33 and Glu34 is contacted by the S1' site (Leu155, Glu157, Met203, Phe206 and His207). **b**, **c**, Lys11 diUb cleavage (**b**) and Lys11 diUb ABP reactivity (**c**) assays with S1' site mutants. **d**, DUB assays with wild-type Cezanne and Ub variants. Lys11 diUb substrates were assembled to specifically mutate the proximal Ub by using K11R, K63R mutations in the distal,

and K63R, Δ LRGG in the proximal Ub moiety. No further mutations were introduced in WT*, while K33A* and K33E* variants additionally contained respective mutations in their proximal Ub only. **e**, Lys11 diUb cleavage assay with Glu157 variants. **f**, **g**, Gel-based specificity analysis of Cezanne E157K. The mutant shows a reduced activity towards Lys11-linked diUb and therefore specificity compared to wild-type Cezanne (compare Fig. 1b). Assays with each variant were performed at least twice with consistent results. For gel source images, see Supplementary Fig. 1.

Extended Data Table 1 | Data collection and refinement statistics

	Cez apo SeMet (88-438)	Cez apo (129-438)	Cez-Ub (129-438, QPG)	Cez-Lys11 diUb (129-438)	A20-Ub (1-366)
Data collection					
Space Group	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 4 ₁ 2 ₁ 2	<i>H</i> 3	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 12 ₁ 1
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	96.72, 96.72, 83.37	103.35, 103.35, 90.20	157.56, 157.56, 75.6	92.03, 56.52, 91.75	64.20, 71.95, 203.93
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 120	90, 90, 90	90, 94.64, 90
Resolution (Å)	50.00 - 3.70 (3.83 - 3.70)	90.20 - 2.20 (2.27 - 2.20)	66.13 - 2.00 (2.05 - 2.00)	56.52 - 2.80 (2.87 - 2.80)	49.33 - 2.85 (2.96 - 2.85)
<i>R</i> _{merge}	0.212 (0.718)	0.071 (0.829)	0.125 (0.854)	0.145 (0.689)	0.070 (0.384)
$\langle I/\sigma \rangle$	11.7 (4.3)	13.1 (1.9)	8.6 (1.9)	9.5 (2.0)	9.6 (2.1)
<i>CC</i> _{1/2}		0.985 (0.437)	0.994 (0.515)	0.990 (0.691)	0.995 (0.867)
Completeness (%)	99.8 (100)	100 (99.9)	100 (100)	94.6 (95.5)	99.4 (99.8)
Redundancy	7.7 (7.6)	7.9 (5.8)	4.2 (4.1)	4.6 (4.6)	3.2 (3.3)
Refinement					
Resolution (Å)		73.08 - 2.20	45.48 - 2.00	48.16 - 2.80	49.33 - 2.85
No. reflections		25351	47215	11478	43393
<i>R</i> _{work} / <i>R</i> _{free}		19.8 / 21.9	17.6 / 21.7	20.7 / 24.4	19.5 / 24.6
No. atoms					
Protein		2143	5599	3318	11558
Ligand/ion		-	18	35	8
Water		90	621	3	-
<i>B</i> factors					
Protein		72.3	29.0	46.8	67.1
Ligand/ion		-	36.9	56.3	67.6
Water		58.8	34.2	17.2	-
R.m.s deviations					
Bond length (Å)		0.002	0.002	0.002	0.002
Bond angles (°)		0.59	0.62	0.52	0.54

Values in parentheses are for the highest resolution shell. All datasets were collected from a single crystal each.

Atomic structure of the entire mammalian mitochondrial complex I

Karol Fiedorczuk^{1,2}, James A. Letts¹, Gianluca Degliesposti³, Karol Kaszuba¹, Mark Skehel³ & Leonid A. Sazanov¹

Mitochondrial complex I (also known as NADH:ubiquinone oxidoreductase) contributes to cellular energy production by transferring electrons from NADH to ubiquinone coupled to proton translocation across the membrane^{1,2}. It is the largest protein assembly of the respiratory chain with a total mass of 970 kilodaltons³. Here we present a nearly complete atomic structure of ovine (*Ovis aries*) mitochondrial complex I at 3.9 Å resolution, solved by cryo-electron microscopy with cross-linking and mass-spectrometry mapping experiments. All 14 conserved core subunits and 31 mitochondria-specific supernumerary subunits are resolved within the L-shaped molecule. The hydrophilic matrix arm comprises flavin mononucleotide and 8 iron–sulfur clusters involved in electron transfer, and the membrane arm contains 78 transmembrane helices, mostly contributed by antiporter-like subunits involved in proton translocation. Supernumerary subunits form an interlinked, stabilizing shell around the conserved core. Tightly bound lipids (including cardiolipins) further stabilize interactions between the hydrophobic subunits. Subunits with possible regulatory roles contain additional cofactors, NADPH and two phosphopantetheine molecules, which are shown to be involved in inter-subunit interactions. We observe two different conformations of the complex, which may be related to the conformationally driven coupling mechanism and to the active–inactive transition of the enzyme. Our structure provides insight into the mechanism, assembly, maturation and dysfunction of mitochondrial complex I, and allows detailed molecular analysis of disease-causing mutations.

The electrochemical proton gradient across the inner mitochondrial membrane required by ATP synthase is maintained by the electron transport chain proton-pumping complexes I, III and IV (refs 1, 2). Complex I is crucial for the entire process, and even mild complex I deficiencies can cause severe pathologies⁴. Mammalian complex I is built of 45 (44 unique) subunits. Fourteen ‘core’ subunits, conserved from bacteria, comprise the ‘minimal’ form of the enzyme^{1,5}, an L-shaped structure with seven subunits in the hydrophilic peripheral arm and another seven in the membrane arm. Mammalian complex I also contains 31 ‘supernumerary’ or ‘accessory’ subunits⁵, forming a shell around the core⁶. The role of these subunits is unclear. Complex I probably translocates four protons for every two electrons transferred from NADH to ubiquinone^{7,8}.

Complex I is the least characterized enzyme of the electron transport chain. The crystal structure of bacterial (*Thermus thermophilus*) complex I is the only full atomic model of the enzyme^{9–11}. In a later structure of the mitochondrial enzyme from aerobic yeast *Yarrowia lipolytica*, the atomic model comprises only about 25% of the protein¹². Studies of bovine complex I resulted in poly-alanine models for the core and 22 supernumerary subunits^{6,13}. Here we present the nearly complete atomic structure of mammalian complex I, containing all subunits and all known cofactors.

We used the ovine (*O. aries*) enzyme (Methods). Classification of cryo-electron microscopy (cryo-EM) images indicated that the relative

orientation between the two arms of the complex is variable, producing classes with either an ‘open’ or ‘closed’ angle between them (Extended Data Fig. 1). Particles in the ‘open’ conformation produced a higher resolution map at ~3.9 Å (Extended Data Fig. 2). The resolution drops at the periphery of the molecule (Extended Data Fig. 3), owing to remaining differences in conformation. Therefore, we performed 3D refinements focusing on the peripheral arm and membrane arm separately, resulting in more uniformly resolved maps for the peripheral arm at 3.9 Å resolution and for the membrane arm at 4.1 Å (Extended Data Fig. 3). The best maps were combined for model building (Fig. 1a, Extended Data Figs 3d and 4).

Modelling of the core subunits was facilitated by the conservation of their fold from bacteria^{6,9}. The assignment of the 31 supernumerary subunits (~0.5 MDa) to the remaining density is challenging. To provide experimental verification for previous assignments, to locate remaining subunits and to obtain restraints on the fold of individual subunits, we performed extensive cross-linking/mass-spectrometry mapping experiments (Extended Data Fig. 5, Supplementary Tables). The initial structure was improved by density-guided re-building in Rosetta¹⁴, resulting in final model of high quality (Extended Data Fig. 2c).

While this manuscript was under review, a 4.2 Å resolution cryo-EM model for the bovine complex I was published¹⁵. The assignments of all subunits agree with our structure, and two major conformations of the complex (somewhat different from ovine) are also observed. However, owing to lower resolution, the completeness of the atomic model is low for the supernumerary subunits (73% of residues are without side chains) and for the core 51-kDa, 24-kDa and 75-kDa subunits (extended data tables 1 and 2 in ref. 15).

In our ovine structure (Fig. 1), subunits were built almost entirely as atomic models with only some surface-exposed loops missing. Subunit B14.7 is disordered, so this area was modelled as poly-alanine according to its clear density in our ovine supercomplex map¹⁶. The model is at the atomic level for 88% of the protein (Extended Data Table 1), presenting, to our knowledge, the most complete atomic structure of mitochondrial complex I so far.

The fold of core subunits is generally conserved from bacteria (Supplementary Discussion). The Fe–S clusters are arranged in the redox chain with distances similar to bovine⁶ and *T. thermophilus*¹⁰ (Fig. 2a). The NADH-binding site is also conserved (Fig. 2b), preserving the entire path for electron transfer from NADH towards quinone. Key features in the membrane domain are also conserved, with four proton channels built around the central axis of polar residues propagating from the quinone-binding (Q) site into the three antiporter-like subunits.

Throughout the article, we use bovine nomenclature with numbering of residues according to mature¹⁷ ovine sequences; see Extended Data Table 1 for human nomenclature. The Q site lies at the interface of the hydrophilic 49-kDa and PSST subunits, and the membrane ND1 and ND3 subunits. The unique structure of the Q site, which forms an enclosed tunnel extending from the membrane towards cluster N2

¹Institute of Science and Technology Austria, Klosterneuburg 3400, Austria. ²MRC Mitochondrial Biology Unit, Cambridge CB2 0XY, UK. ³MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.

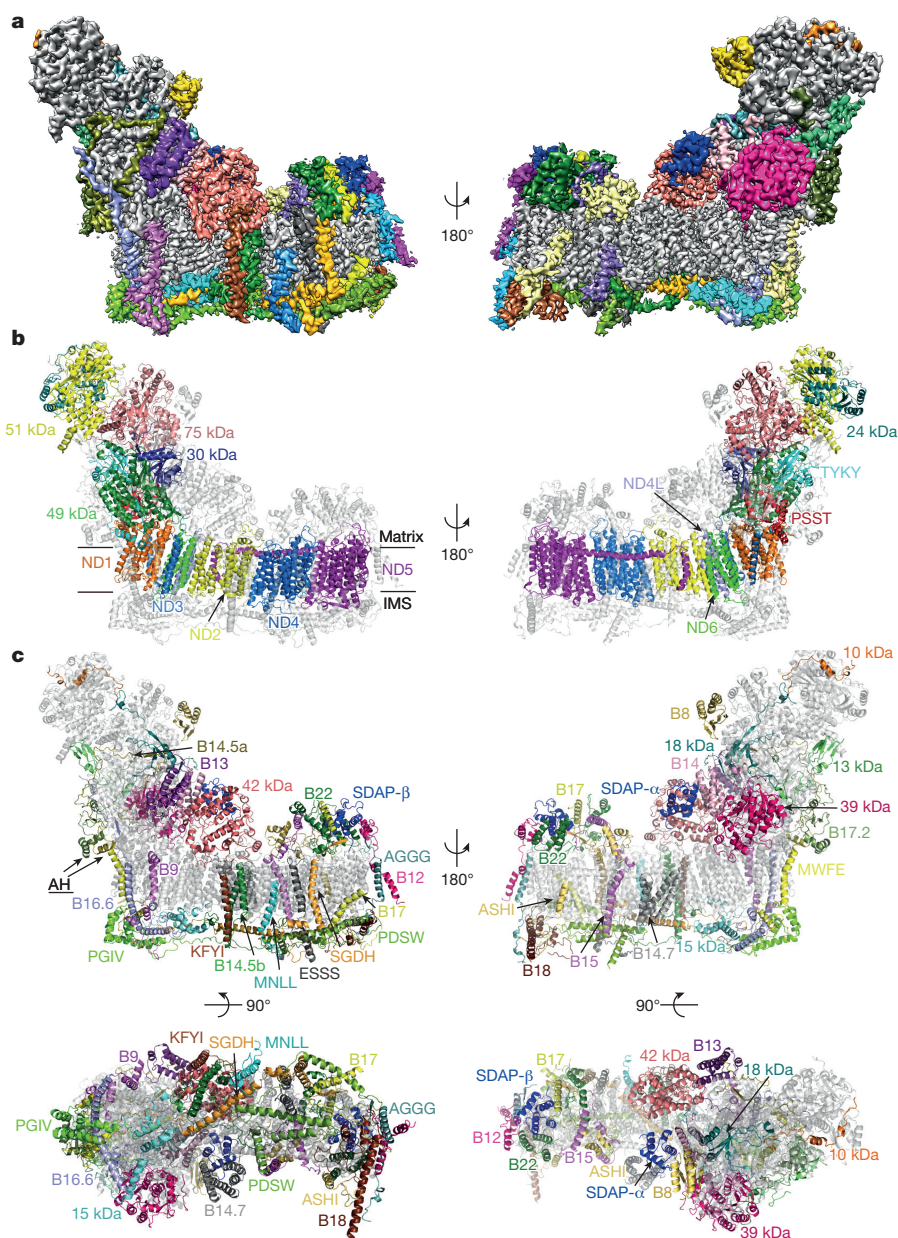


Figure 1 | Structure of ovine complex I. a, Cryo-EM density coloured by subunit, with core subunits in grey (left–right view). **b,** Structure depicted as a cartoon, with core subunits coloured and labelled, and supernumerary subunits in grey and transparent. Approximate lipid bilayer boundaries

about 25 Å away, is conserved with one difference: a loop connecting two strands of the N-terminal β -sheet from the 49-kDa subunit ($\beta 1$ – $\beta 2^{49\text{-kDa}}$ loop) extends further into the cavity, clashing with the position of the bound quinone from the bacterial structure⁹, where it interacts with conserved His59^{49-kDa} and Tyr108^{49-kDa} (Fig. 2c, d). A similar conformation was observed in the yeast enzyme, leading to the proposal that it represents the ‘deactive’ state¹². In the absence of substrates, mitochondrial complex I exists in the deactive state (which may prevent oxygen radical production *in vivo*¹⁸), and converts into the ‘active’ state¹⁹ only upon turnover. Because the $\beta 1$ – $\beta 2^{49\text{-kDa}}$ loop in our structure will prevent quinone access closer than 20 Å to cluster N2 (blocking electron transfer), it probably also represents the deactive state. The ‘closed’ class conformation resembles one in supercomplex¹⁶, so may be more physiological. It remains to be established whether, as discussed previously^{12,15}, different observed conformations are related to the catalytic cycle or indeed to active/deactive transitions, but the overall conformational flexibility of the complex is clear.

are indicated. **c**, Structure depicted with core subunits in grey and supernumerary subunits coloured and labelled (left-right, IMS-matrix views). Amphipathic helices at the 'heel' of the complex, probably attached to the lipid bilayer, are indicated as AH.

Supernumerary subunits form a shell around the core subunits^{6,12}, especially around the membrane domain and its interface with the peripheral arm. With few exceptions, most supernumerary subunits are not globular, but form extended structures containing α -helices and coils (Extended Data Fig. 6), allowing for numerous interactions at interfaces with other subunits (Extended Data Table 2). They interweave extensively with each other and the core subunits (Extended Data Fig. 7), making the whole mitochondrial complex assembly much more interlinked and thus more stable, with a large total buried surface area (Extended Data Table 2). The intertwined nature of subunit structures suggests that they can be added to the complex only in a certain order, and, therefore, that the assembly of subunits must be tightly controlled²⁰.

The fold of supernumerary subunits is described in the Supplementary Discussion. In summary, those associated with the membrane arm include 12 single transmembrane helix domain²¹ subunits scattered around the entire arm. Six of these surround the

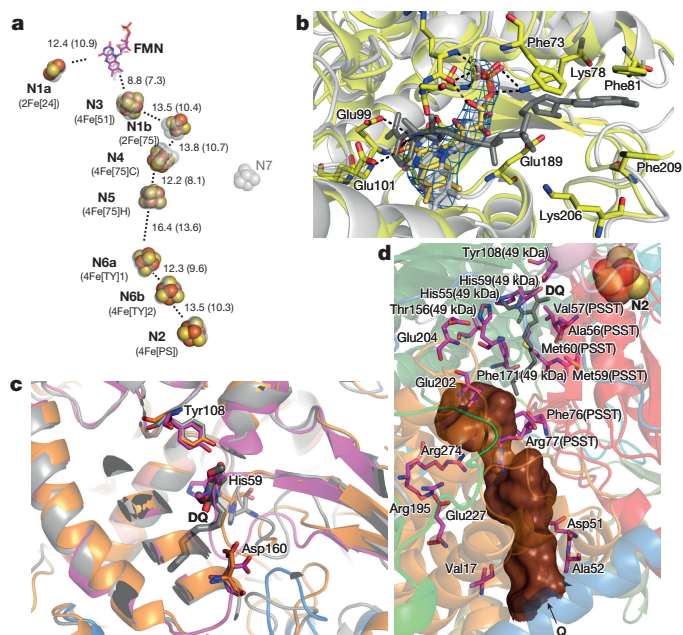


Figure 2 | Arrangement of redox centres and substrate binding sites. **a**, Fe–S clusters are shown as spheres with centre-to-centre and edge-to-edge (in brackets) distances indicated in Å, overlaid with transparent grey depictions from *T. thermophilus*. Both traditional and structure-based (in brackets) nomenclature for clusters is shown. **b**, NADH-binding site (overlay with *T. thermophilus* structure in grey, containing NADH). Cryo-EM density for flavin mononucleotide (FMN) is shown in blue. Key residues involved in interactions with FMN and NADH are shown as sticks. **c**, Quinone-binding site with subunits coloured as Fig. 1. Key 31–32⁴⁹-kDa loop deviates from bacterial structure (grey) and is more similar to *Y. lipolytica* (orange, PDB 4WZ7; ref. 12), clashing with the decyl-ubiquinone (DQ) head group position in *T. thermophilus* (grey; ref. 9). **d**, Environment surrounding the Q cavity (brown surface, entrance point indicated by an arrow), with some of the functionally important residues shown as sticks and labelled with non-ND1 subunit names in brackets. The quinone from the aligned *T. thermophilus* structure is shown in grey (DQ), demonstrating that the distal part of the cavity is blocked in the ovine enzyme.

membrane arm tip and contribute their intertwined N-terminal domains to a large matrix ‘bulge’, the bulk of which is formed by an acyl-carrier protein (ACP)–LYR motif subunit pair (SDAP-β–B22). The large globular 42-kDa subunit from the nucleoside kinase family is attached to the matrix side of ND2 near the peripheral arm interface. On the intermembrane space (IMS) side, subunits SGD and PDSW are ‘interlocked’ via their backbone and contain three long α-helices traversing nearly the entire membrane arm (Fig. 1c). PDSW and the subunits with CHCH domains (PGIV, 15 kDa and B18) contain disulfide bonds that further stabilize the fold in the oxidizing environment of the IMS. PGIV clamps the ‘heel’ of the complex to the middle of the membrane arm. The disulfide-rich, interlocked helices of the IMS subunits, with their rigid and stable structure, appear to replace the hairpin/helix motif (βH) found in bacterial complex I (refs 9, 11).

Subunits associated with the peripheral arm include the NADPH-containing 39-kDa subunit, the Zn-containing 13-kDa subunit and another ACP–LYR motif pair, SDAP-α–B14, with the latter pair and B13 jointly ‘embracing’ the 42-kDa subunit. The interface between the peripheral arm and membrane domain is stabilized by the exceptionally long membrane-traversing helix of subunit B16.6, as well as by B17.2 and B14.5a, both of which contain N-terminal amphipathic α-helices bound at the membrane interface, with the rest of their polypeptides wrapping around the hydrophilic arm. Subunits PSST, TYKY and B9 also contain such amphipathic helices, all located at the heel of the complex (Fig. 1c), probably assisting in proper peripheral arm position over the lipid bilayer.

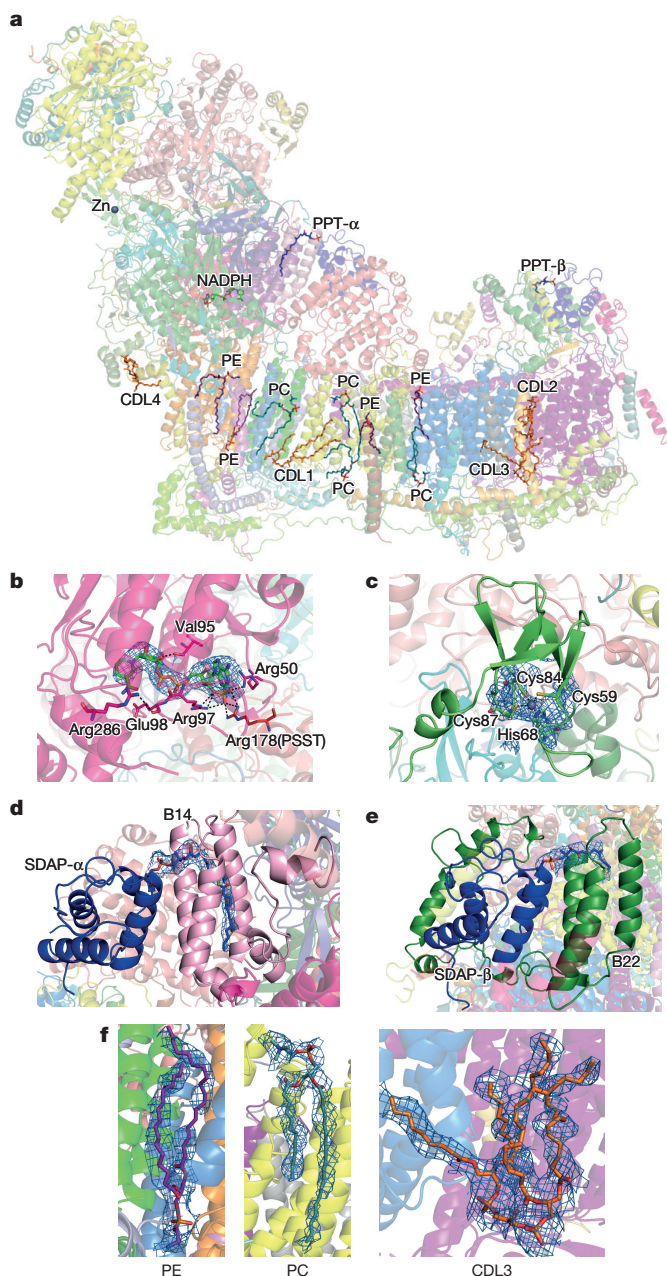


Figure 3 | Additional cofactors identified in the structure. **a**, Overview of the model, coloured as in Fig. 1c, with cofactors shown as sticks. CDL, cardiolipin; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PPT, phosphopantetheine. **b**, NADPH in the 39-kDa subunit. Interacting residues are shown. **c**, Zn²⁺ ion in the 13-kDa subunit, with coordinating residues. **d**, Phosphopantetheine in SDAP-α. **e**, Phosphopantetheine in SDAP-β. **f**, Lipids phosphatidylethanolamine, phosphatidylcholine and cardiolipin. All cofactors are shown with cryo-EM density carved to within 5 Å.

Several cofactors present in supernumerary subunits are well resolved in the structure (Fig. 3). The 39-kDa subunit is wedged into the side of the peripheral arm near the membrane arm interface. It contains a tightly bound non-catalytic NADPH (Fig. 3b) that interacts with conserved Arg178^{PSST}, providing a possible mitochondria redox state-sensitive conformational link to cluster N2. In SDAP-α, a phosphopantetheine that is covalently linked to Ser44 extends its attached acyl chain in the flipped-out²² conformation into the hydrophobic crevice between the helices of the LYR subunit B14 (Fig. 3d). A similar interaction is observed in the SDAP-β–B22 pair. These are the first structures of ACP–LYR complexes showing that their interaction depends on the extended acyl chain and revealing the role of LYR motif.

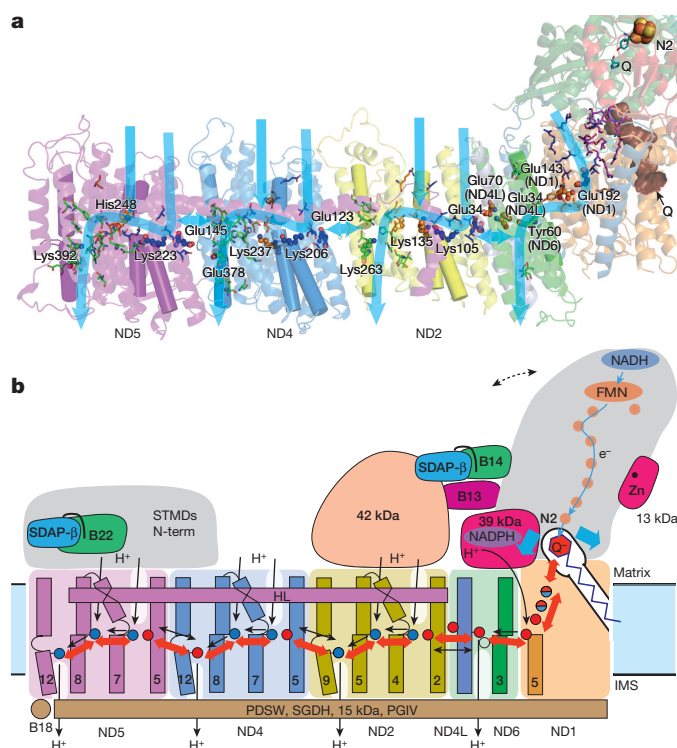


Figure 4 | Mechanism of mitochondrial complex I. **a**, Structure of the core subunits of ovine complex I, coloured as in Fig. 1b, with polar residues in proton channels shown as sticks, with carbon in blue, orange and green for input, connecting and output parts, respectively. Key residues, Glu (TM5), Lys (TM7), Lys/His (TM8) and Lys/Glu (TM12) from the antiporters and the corresponding residues in the E-channel (near Q site), are shown as small spheres and labelled. These residues sit on flexible loops in discontinuous transmembrane helices shown as cylinders. Polar residues linking the E-channel to the Q cavity (brown) are shown in magenta. Tyr108^{49-kDa} and His59^{49-kDa} are shown in cyan near the position of bound Q in bacteria. Possible proton translocation pathways are indicated by blue arrows. **b**, Graphic of the coupling mechanism. Core and some putatively regulatory supernumerary subunits are shown. Conformational changes, indicated by red arrows, propagate from the Q site/E-channel to antiporter-like subunits via the central hydrophilic axis. Shifts of helices near the cluster N2 (ref. 31; blue arrows) may help initiate the process. ND5 helix HL and traverse helices from four supernumerary subunits on the IMS side may serve as stators. Dashed line indicates the shift of peripheral arm in the closed conformation (Extended Data Fig. 8). The NADPH-containing 39-kDa subunit and Zn-containing 13-kDa subunit are essential for activity and may serve as redox sensors. Both SDAP subunits interact with their LYR partners via flipped-out phosphopantetheine (black line). The net result of one conformational cycle, driven by NADH:ubiquinone oxidoreduction, is the translocation of four protons across the membrane (black lines indicate possible pathways).

Complex I is active only when fully assembled with the SDAP- α -B14 pair²³. This interaction, which depends on the acyl chain attached to the ACP, may provide a regulatory link between fatty acid synthesis and oxidative phosphorylation in mitochondria. The 13-kDa subunit contains a Zn-binding motif, coordinating a Zn²⁺ ion in the vicinity of clusters N6a and N5. Zn-containing proteins are sensitive to oxidative stress²⁴, and loss of the 13-kDa subunit leads to loss of cluster N6a (ref. 25) as it becomes exposed. In this way, complex I may be equipped with an oxidative-stress 'sensor', in addition to bound NADPH.

Twelve bound lipids were identified in crevices between hydrophobic subunits. Several observed lipid molecules have four acyl chains and were therefore assigned as cardiolipins, known to be essential for activity²⁶. Notably, a cardiolipin (CDL1; Fig. 3a) and three other lipid molecules fill the void left by the missing (in metazoans) three ND2 N-terminal helices⁶. This void is encircled by the two-transmembrane

helix subunit B14.5b and the single transmembrane helix domain subunit KFYI, indicating that the ND2 helices may have been lost in evolution to accommodate a specific binding site for lipids. Two cardiolipins (CDL2 and CDL3) fill a large gap between the antiporter-like ND4 and ND5 subunits, preventing potential proton leaks and instability. Another cardiolipin (CDL4) stabilizes amphipathic helices at the heel of the complex. The structure thus shows the basis for the essential role of cardiolipin and other lipids.

The mechanism of coupling between electron transfer and proton translocation is still enigmatic. Conservation of key features from bacteria to mammals suggests that the basic mechanism is probably the same, with add-on 'stabilizers' and 'regulators'. As we proposed previously⁹, the central axis of polar residues in the membrane probably has a key role (Fig. 4a). In each catalytic cycle, the negative charge stored either on Q or on nearby residues in the enclosed Q site may drive conformational changes in ND1 and the proton channel near the Q site, which would propagate via the central axis to channels in antiporter-like subunits ND2, ND4 and ND5, resulting in changes in pK_a and accessibility of key residues. The net result would be the pumping of four protons per cycle, one per each channel. The observed conformation of loops in the Q site probably reflects the deactive state. This conformation might also occur during normal function when quinol is ejected from the site into the lipid bilayer, if active/deactive transitions are related to conformations encountered during the catalytic cycle¹². Supernumerary subunits implicated in active/deactive transitions (39 kDa, B13 and SDAP- α -B14 pair) could also participate in catalytic conformational changes by interacting with the key TM1–TM2^{ND3} loop flanking the Q site, and possibly through interactions with the 42-kDa subunit (Fig. 4b). In the 'closed' class, B13 and SDAP- α move towards the 42-kDa subunit (Extended Data Fig. 8), hinting at such a possibility. Because the 42-kDa subunit is metazoan-specific²⁷, its role may be to fine tune movements during turnover. The traverse helix HL from ND5 appears to mainly have a stabilizing 'stator' role²⁸ rather than being a moving element²⁹. Rigid disulfide-rich supernumerary subunits traversing the IMS side of the membrane domain may represent another stator element unique to the mitochondrial enzyme (Fig. 4b).

Our structure clearly shows that supernumerary subunits stabilize the complex. Some of them, especially those containing additional cofactors (39 kDa, SDAPs, B14, B22 and 13 kDa) and phosphorylated residues (42 kDa, ESSS, MWFE, B14.5a, B14.5b and B16.6)³⁰, may provide regulatory links to the redox status of the cell, lipid biosynthesis and mitochondrial homeostasis. Known human pathological mutations are present in all of the core and many of the supernumerary subunits⁴. Our structure provides the framework for understanding the molecular basis of mutations and mechanisms of complex I function and regulation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 August; accepted 26 August 2016.

Published online 5 September 2016.

1. Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388 (2015).
2. Moser, C. C., Farid, T. A., Chobot, S. E. & Dutton, P. L. Electron tunneling chains of mitochondria. *Biochim. Biophys. Acta* **1757**, 1096–1109 (2006).
3. Letts, J. A. & Sazanov, L. A. Gaining mass: the structure of respiratory complex I from bacterial towards mitochondrial versions. *Curr. Opin. Struct. Biol.* **33**, 135–145 (2015).
4. Fassone, E. & Rahman, S. Complex I deficiency: clinical features, biochemistry and molecular genetics. *J. Med. Genet.* **49**, 578–590 (2012).
5. Walker, J. E. The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. *Q. Rev. Biophys.* **25**, 253–324 (1992).
6. Vinothkumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* **515**, 80–84 (2014).
7. Galkin, A., Dröse, S. & Brandt, U. The proton pumping stoichiometry of purified mitochondrial complex I reconstituted into proteoliposomes. *Biochim. Biophys. Acta* **1757**, 1575–1581 (2006).

8. Wikström, M., Sharma, V., Kaila, V. R., Hosler, J. P. & Hummer, G. New perspectives on proton pumping in cellular respiration. *Chem. Rev.* **115**, 2196–2221 (2015).
9. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
10. Sazanov, L. A. & Hinchliffe, P. Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* **311**, 1430–1436 (2006).
11. Efremov, R. G. & Sazanov, L. A. Structure of the membrane domain of respiratory complex I. *Nature* **476**, 414–420 (2011).
12. Zickermann, V. *et al.* Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347**, 44–49 (2015).
13. Zhu, J. *et al.* Structure of subcomplex I β of mammalian respiratory complex I leads to new supernumerary subunit assignments. *Proc. Natl Acad. Sci. USA* **112**, 12087–12092 (2015).
14. DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015).
15. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536**, 354–358 (2016).
16. Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. Architecture of respiratory supercomplexes. *Nature* <http://dx.doi.org/10.1038/nature19774> (2016).
17. Carroll, J., Fearnley, I. M., Shannon, R. J., Hirst, J. & Walker, J. E. Analysis of the subunit composition of complex I from bovine heart mitochondria. *Mol. Cell. Proteomics* **2**, 117–126 (2003).
18. Chouchani, E. T. *et al.* Cardioprotection by S-nitrosation of a cysteine switch on mitochondrial complex I. *Nat. Med.* **19**, 753–759 (2013).
19. Vinogradov, A. D. & Grivennikova, V. G. The mitochondrial complex I: progress in understanding of catalytic properties. *IUBMB Life* **52**, 129–134 (2001).
20. Sánchez-Caballero, L., Guerrero-Castillo, S. & Nijtmans, L. Unraveling the complexity of mitochondrial complex I assembly: A dynamic process. *Biochim. Biophys. Acta* **1857**, 980–990 (2016).
21. Zickermann, V., Angerer, H., Ding, M. G., Nübel, E. & Brandt, U. Small single transmembrane domain (STMD) proteins organize the hydrophobic subunits of large membrane protein complexes. *FEBS Lett.* **584**, 2516–2525 (2010).
22. Cronan, J. E. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. *Biochem. J.* **460**, 157–163 (2014).
23. Angerer, H. Eukaryotic LYR proteins interact with mitochondrial protein complexes. *Biology (Basel)* **4**, 133–150 (2015).
24. Oteiza, P. I. Zinc and the modulation of redox homeostasis. *Free Radic. Biol. Med.* **53**, 1748–1759 (2012).
25. Kmita, K. *et al.* Accessory NUMM (NDUFS6) subunit harbors a Zn-binding site and is essential for biogenesis of mitochondrial complex I. *Proc. Natl Acad. Sci. USA* **112**, 5685–5690 (2015).
26. Sharpley, M. S., Shannon, R. J., Draghi, F. & Hirst, J. Interactions between phospholipids and NADH:ubiquinone oxidoreductase (complex I) from bovine mitochondria. *Biochemistry* **45**, 241–248 (2006).
27. Brandt, U. Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* **75**, 69–92 (2006).
28. Belevich, G., Knuuti, J., Verkhovsky, M. I., Wikström, M. & Verkhovskaya, M. Probing the mechanistic role of the long α -helix in subunit L of respiratory Complex I from *Escherichia coli* by site-directed mutagenesis. *Mol. Microbiol.* **82**, 1086–1095 (2011).
29. Efremov, R. G., Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465**, 441–445 (2010).
30. Palmisano, G., Sardanelli, A. M., Signorile, A., Papa, S. & Larsen, M. R. The phosphorylation pattern of bovine heart complex I subunits. *Proteomics* **7**, 1575–1583 (2007).
31. Berrisford, J. M. & Sazanov, L. A. Structural basis for the mechanism of respiratory complex I. *J. Biol. Chem.* **284**, 29773–29783 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the ETH Zurich ScopeM Center for access to Titan Krios EM. Data processing was performed at the IST high-performance computer cluster. K.F. is partially funded by the Medical Research Council UK PhD fellowship. J.A.L. holds a long-term fellowship from FEBS. This project has received funding from the European Union's 2020 research and innovation programme under grant agreement No 701309.

Author Contributions K.F. purified complex I for grid preparation, prepared cryo-EM grids, acquired and processed EM data, and co-built the models; J.A.L. purified complex I for cross-linking experiments, analysed cross-linking data and co-built the models; G.D. performed cross-linking/mass-spectrometry experiments, K.K. performed model re-building in Rosetta and sequence alignments; G.D. and M.S. analysed cross-linking data; L.A.S. designed and supervised the project, processed and analysed data and wrote the manuscript, with contributions from all authors.

Author Information The electron microscopy maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-4084, EMD-4090, EMD-4091 and EMD-4093 (combined map). The model has been deposited in the Protein Data Bank (PDB) under accession code 5LNK. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.S. (sazanov@ist.ac.at).

Reviewer Information *Nature* thanks P. Ädelroth, M. Ryan and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein purification and electron microscopy. Protein was purified from *O. aries* heart mitochondria following the protocol adapted with some modifications from a previously published procedure for the bovine enzyme (ref. 32 and J.A.L. *et al.*, manuscript submitted). We explored *O. aries* as a source of complex I that may be more suitable for high-resolution structural studies than the extensively studied bovine enzyme. We find that ovine enzyme appears more stable, as it is highly active after purification and retains the 42-kDa subunit, easily lost from bovine complex (J.A.L. *et al.*, manuscript submitted). In terms of overall sequence similarity ovine is as good a model of the human enzyme as bovine (~84%), and all 44 different subunits of complex I were identified in the preparation by mass spectrometry (J.A.L. *et al.*, manuscript submitted). In brief, fresh ovine hearts were purchased from the local abattoir and mitochondria prepared as described previously³². Mitochondrial membranes were solubilized in the branched chain detergent lauryl maltose neopentyl glycol (LMNG, 1%) and the sample applied to Q-sepharose HP anion exchange column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.4, 10% (v/v) glycerol, 1 mM EDTA, 1 mM DTT and 0.1% LMNG. Protein was eluted with a NaCl gradient, peak fractions concentrated and applied to Superose 6 HiLoad 16/60 column equilibrated in 20 mM HEPES, pH 7.4, 2 mM EDTA, 1.5% (v/v) glycerol, 100 mM NaCl and 0.02% Brij-35. The peak fraction was concentrated to ~5 mg ml⁻¹ protein and ~0.2% Brij-35. Then, 2.7 µl of sample was applied to glow discharged Quantifoil R 0.6/1 copper grids and blotted for 34 s at 90% humidity in the chamber of FEI Vitrobot III. Immediately after, the sample was snap-frozen in liquid ethane. Extensive trials with different detergents, including previously used Cymal-7 (ref. 6), revealed Brij-35 as the detergent giving the most homogeneous spread of particles. Imaging was performed with a 300 kV Titan Krios electron microscope equipped with direct electron detector FEI Falcon-II (ETH Zurich, ScopeM centre) in automated data collection mode at a calibrated magnification of 1.39 Å pixel⁻¹ (×100,720) and dose of 26 e s⁻¹ Å⁻² with total 3-s exposure time. The data were collected as seven movie frames fractionated over the first second of exposure and an averaged image over 3 s.

Image processing. We collected a total of 2.6 k micrographs in two datasets, which were combined. All processing steps were done using RELION³³ unless otherwise stated. We used averaged images from high dose 3 s exposure for initial CTF estimations using CTFFIND4 (ref. 34) and for automated particle picking in Relion, resulting in ~241 k particles. MOTIONCORR³⁵ was used for whole-image drift correction of movie frames 1–7 (1 s) of each micrograph. Contrast transfer function (CTF) parameters of the corrected micrographs were estimated using Gctf and refined locally for each particle³⁶. The particles were extracted using 296² pixel box and sorted by reference-free 2D classification, resulting in ~171 k particles selected from good 2D classes. These were used for 3D classification with a regularization parameter T of 8 and a 30 Å low-pass filtered initial model from a previous low resolution model of the bovine enzyme⁶. That resulted in ~130 k particles of good quality; however, it was clear that the relative orientation between the two arms of the complex is slightly variable, producing 3D classes with either an open or closed angle between the arms (Extended Data Fig. 1). Particles in the open conformation (~82 k particles) produced higher resolution maps and were selected for a final reconstruction. For all high-resolution refinements, particles were re-extracted from the motion corrected micrographs with a 512² pixel box to allow for high-resolution CTF correction³⁷. After initial auto-refinement, particle-based beam-induced motion correction and radiation-damage weighing (particle polishing) was performed³⁸. Refinement of polished particles gave a map resolved to 3.9 Å. All resolutions are based on the gold-standard (two halves of data refined independently) FSC = 0.143 criterion³⁹. This 3D class selection probably still allows for small variations in the conformation, therefore the local resolution varies within the map, especially at the extremities of both arms (Extended Data Fig. 3). At the periphery of the molecule the resolution drops not only owing to the usual decrease in the precision of particle alignments in these areas, but also due to differences in the protein conformation, greatest at the edges of the molecule. To overcome this limitation we performed 3D refinement focused on the peripheral and membrane domains separately (with the subtraction of signal from the remaining parts of the complex⁴⁰). This resulted in a 3.9 Å map of the peripheral arm, very well resolved in all areas, including the edges of the domain. The membrane domain refined to 4.1 Å, however, the map was more uniform and so better resolved for the distal part of the domain (near subunit ND5) as compared to the density from the refinement of the entire complex (Extended Data Fig. 3). Higher quality refinement of the peripheral arm probably stems from the fact that high electron density of eight Fe-S clusters helps in particle alignment. The least ordered part of the complex is the 42-kDa subunit, loosely attached to the membrane domain. We have performed extended 3D classification of the

open class to identify the most homogeneous population, especially with respect to 42-kDa subunit. This class (64 k particles) was refined to 4.0 Å, and the resulting density was used to model the 42-kDa subunit. To assist with overall model building and refinement, several maps were carved around specific parts of the complex and combined into one map in UCSF Chimera⁴¹: peripheral arm from peripheral-arm-focused refinement, the area around ND4/5 subunits (tip of the membrane domain) from membrane-arm-focused refinement, the 42-kDa subunit density as above, and the rest of the complex from the overall 3.9 Å map for the open class (Fig. 1a and Extended Data Fig. 3). The final model was refined against this combined map.

The final map is of high quality, with about three-quarters of the map at 3.9 Å resolution and the rest at 4.1 Å. Large- and medium-size side chains, as well as relatively small Val and Thr, are clearly seen in the density (Extended Data Fig. 4). Carboxylates (Asp, Glu) have much lower density than other residues owing to early radiation damage, as observed previously⁴². Disulfide bridges also are subject to early damage, as in X-ray crystallography⁴³. Few features at the interfaces of maps used for the combined map may be better resolved in individual maps, since in overlapping regions both maps contribute. For example, the β1–β2⁴⁹-kDa loop is better resolved in the peripheral-arm-focused map, which is deposited along with other constituent maps. Overall map filtered to lower resolution is very similar to the previous 5 Å resolution map for the bovine enzyme⁶, suggesting that the mammalian complex I structure is very well conserved. One difference is that in ovine complex the accessory four-transmembrane subunit B14.7 is disordered in the detergent used for the microscopy samples (Brij-35). It is likely to be disordered rather than detached as B14.7 was identified by mass spectrometry in the sample used for electron microscopy (data not shown). Since Brij-35 gave us the best yield of particles, we kept its use for data collection, but took advantage of the availability of cryo-EM maps of ovine respiratory supercomplexes in our laboratory. In these maps all the subunits of complex I are well ordered, and so in our final complex I model we included the poly-ALA model of B14.7 based on 5.8 Å resolution map of the ‘tight’ respirasome¹⁶. Loss of B14.7 also results in the disorder of the nearby C-terminal half of transverse helix HL and TM16 from ND5, as well as TM4 from ND6, which were also modelled as poly-alanine (these stretches can be recognized by B-factor set to 200) based on the tight respirasome map. The register in poly-alanine stretches is approximate. The density for the 42-kDa subunit is rather weak but this subunit clearly preserves the nucleoside kinase family fold, which allowed us to model most of it using Rosetta and visible large side-chains as a guide.

Model building and refinement. For the 14 core subunits the initial homology models were generated manually based on the *T. thermophilus* structure⁹ with side-chains rebuilt to ovine sequence using SQWRLA software⁴⁴. Homology models were generated with Phyre2 (ref. 45) and Swiss-model⁴⁶ servers for all supernumerary subunits, although they were mostly useful only for subunits with large globular domains, such as the 42-kDa and 39-kDa subunits, as well as for those with known structure of close homologues (SDAPs and B8). Secondary structure predictions for all subunits were generated with PredictProtein⁴⁷, PsiPred⁴⁸ and TMMHMM⁴⁹ servers, and were helpful during model building. Initial assignments of the location and the fold of supernumerary subunits were based on our cross-linking data and the secondary structure features and side-chain density observed in the cryo-EM map, with checks for consistency with the knowledge on subcomplexes and assembly intermediates in complex I. The initial models were adjusted to cryo-EM density (in cases when homology models were useful) or built manually in COOT⁵⁰. Lipids were tentatively assigned on the basis of appearance in the density as cardiolipins, phosphatidylcholines and phosphatidylethanolamines, known to co-purify with the complex^{26,51}.

Initial models were re-built and refined in Rosetta release version 2016.02.58402 using protocols optimized for cryo-EM maps¹⁴. For each subunit, 100 different models were produced in Rosetta with optimization of density fit using elec_dens_fast function (with -denswt = 40, chosen from several trials), selection of the best fitting structure and structure relaxation using -FastRelax flag. From the produced structures several best-scoring by density fit and geometry were selected and used in COOT to guide further model building/optimization. This procedure resulted mainly in improvements to backbone geometry, especially in coils, still allowing for the good fit of side-chains into density.

After several rounds of re-building the final model was refined with the Phenix suite⁵² phenix.real_space_refine program for 5 macro-cycles using the electron scattering table with default and secondary structure restraints. This resulted in a high quality model in terms of geometry (Molprobit score 2.5, that is, corresponding to average structure at 2.5 Å resolution) and fit to density (Extended Data Fig. 2c). **Cross-linking.** All the cross-linking reactions were performed using purified solutions of complex I at a concentration of 1 mg ml⁻¹. Following experimental optimisation, ten separate experiments were performed. Experiments varied in relation to the detergent added to the buffer (DDM, LMNG or LDAO/DDM), the cross-linking reagent (targeting lysine or acidic residues) and the protease used to digest the samples (trypsin or endoproteinase Glu-C) (Supplementary Table 1).

Isotopically labelled cross-linking reagents were purchased from Creative Molecules (Canada). 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) was purchased from Sigma. Homobifunctional, isotopically-coded *N*-hydroxysuccinimide (NHS) esters disuccinimidyl suberate (DSS H₁₂/D₁₂), bis-sulfodisuccinimidyl suberate (BS3 H₁₂/D₁₂) and disuccinimidyl adipate (DSA ¹²C₆/¹³C₆) were used at a final concentration of 50 µM as cross-linking reagents to target lysine residues. The reactions were incubated for 45 min at 37 °C and quenched by adding NH₄HCO₃ to a final concentration of 50 mM and incubating for further 15 min. Isotopically labelled adipic acid dihydrazide (ADH H₈/D₈) and suberic acid dihydrazide (SDH H₁₂/D₁₂) were used to target the acidic residues, using DMTMM as catalyst. The cross-linking reaction was initiated by adding ADH or SDH and DMTMM to final concentrations 5 mg ml⁻¹, 6 mg ml⁻¹ and 12 mg ml⁻¹, respectively. The samples were incubated at 37 °C for 60 min and the reactions stopped using gel filtration (Zeba Spin Desalting columns 7K MWCO).

The cross-linked samples were freeze-dried and then resuspended in 50 mM NH₄HCO₃, 8 M urea and 0.1% SDS to a final concentration of 1 mg ml⁻¹. Size exclusion protein fractionation was performed through a Superdex 200 Increase 3.2/300 column (GE Healthcare) with 50 mM NH₄HCO₃, 8 M urea and 0.1% SDS as mobile phase at a flow rate of 25 µl min⁻¹. Two-minute fractions were collected and their protein content evaluated by SDS-PAGE. Fractions of similar content were pooled into 4–5 main fractions and concentrated to 1 mg ml⁻¹ using Amicon Ultra-0.5 mL Centrifugal Filters (Millipore).

The filtered cross-linked samples were then enzymatically digested. Samples were freeze-dried and resuspended in 50 mM NH₄HCO₃ and 8 M urea to a final protein concentration of 1 mg ml⁻¹, reduced with 10 mM DTT and alkylated with 50 mM iodoacetamide. Following alkylation, samples were diluted with 50 mM NH₄HCO₃ to 1 M urea before trypsin digestion (or 2 M for Glu-C digestion). Trypsin and Glu-C were added at an enzyme-to-substrate ratio of 1:20 and 1:100, respectively. Digestions were carried out overnight at 37 °C and 25 °C for trypsin and Glu-C respectively. After digestion, the samples were acidified with formic acid to a final concentration of 2% (v/v) and the peptides fractionated by peptide size exclusion chromatography, using a Superdex Peptide 3.2/300 (GE Healthcare) with 30% (v/v) acetonitrile/0.1% (v/v) TFA as mobile phase and at a flow rate of 50 µl min⁻¹. Fractions were collected every 2 min over the elution volume 1.0 ml to 1.7 ml. Before LC-MS/MS analysis fractions were freeze dried and resuspended in 2% (v/v) acetonitrile and 2% (v/v) formic acid.

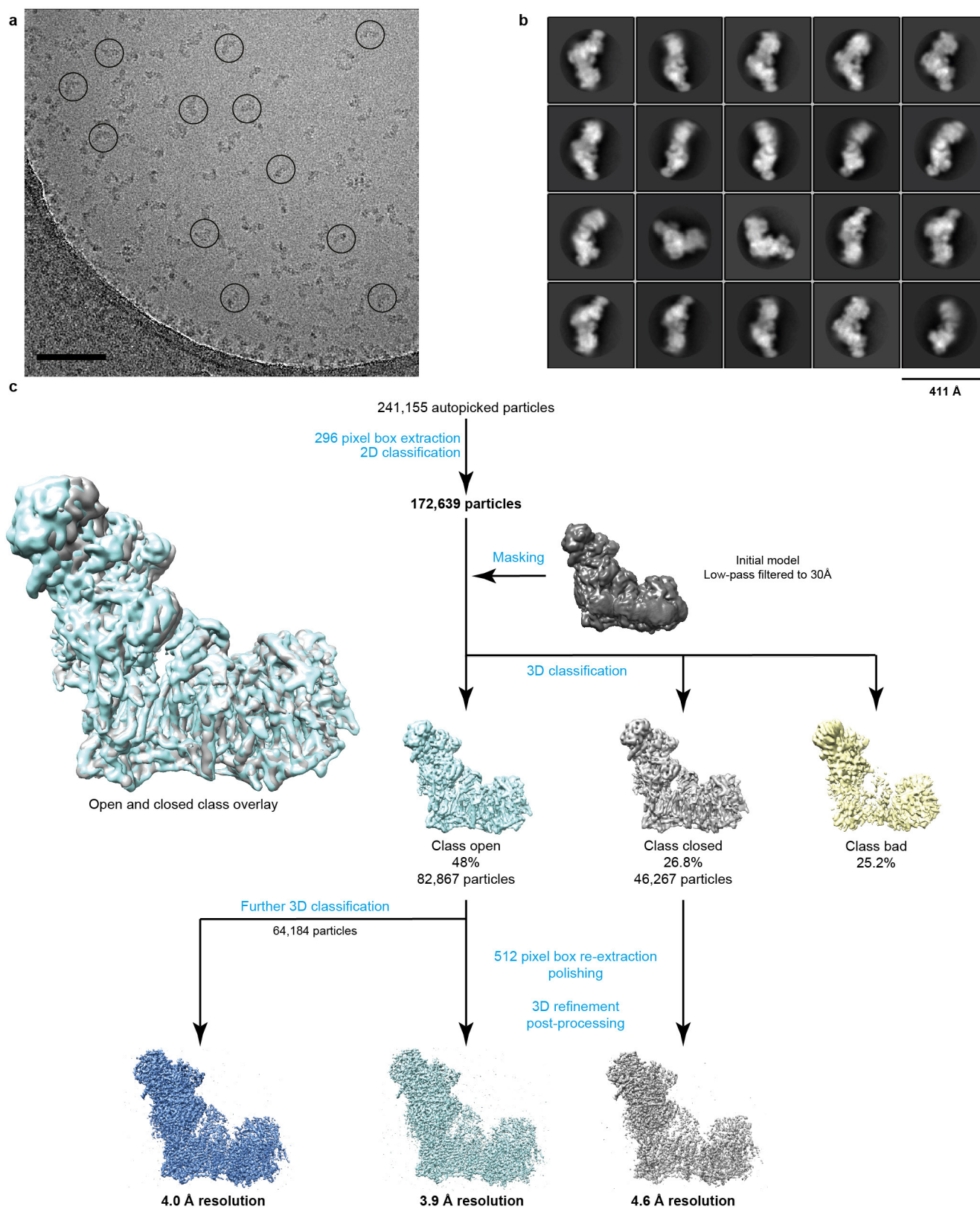
The digests were analysed by nano-scale capillary LC-MS/MS using an Ultimate U3000 HPLC (ThermoScientific Dionex) to deliver a flow of approximately 300 nl min⁻¹. A C18 Acclaim PepMap100 5 µm, 100 µm × 20 mm nano-Viper (ThermoScientific Dionex), trapped the peptides before separation on a C18 Acclaim PepMap100 3 µm, 75 µm × 250 mm nanoViper (ThermoScientific Dionex). Peptides were eluted with a gradient of acetonitrile. The analytical column outlet was directly interfaced via a nano-flow electrospray ionisation source, with a hybrid dual pressure linear ion trap mass spectrometer (Orbitrap Velos, ThermoScientific). Data-dependent analysis was carried out, using a resolution of 30,000 for the full mass spectrometry spectrum, followed by ten MS/MS spectra in the linear ion trap. Mass spectrometry spectra were collected over a *m/z* range of 300–2000. MS/MS scans were collected using threshold energy of 35 for collision-induced dissociation.

For data analysis, Xcalibur raw files were converted into the open mzXML format through MSConvert (Proteowizard) with a 32-bit precision. mzXML files were directly used as input for xQuest searches on a local xQuest installation⁵³. The selection of cross-linked precursor MS/MS data was based on the following criteria: a mass difference among the heavy and the light cross-linker of: 12.07532 Da for BS3, DSS and SDH, 6.02016 Da for DSA and 8.05016 Da for ADH; precursor charge ranging from 3+ to 8+; maximum retention time difference 2.5 min. Searches were performed against an ad hoc database containing all the sequences of ovine complex I subunits together with their reverse used as decoy database. The following parameters were set for xQuest searches: maximum number of missed cleavages (excluding the cross-linking site) 3; peptide length 4–50 amino acids; fixed modifications carbamidomethyl-Cys (mass shift 57.02146 Da); mass shift of the light cross-linker 138.06808 Da for DSS and BS3, 138.0906 Da for ADH, 110.03675 for DSA and 166.1218 for SDH; mass shift of mono-links 156.0786 and 155.0964 Da for DSS and BS3, 138.0906 Da for ADH, 127.0628 Da and 128.0468 Da for DSA, and 184.1324 Da for SDH; MS1 tolerance 10 ppm, MS2 tolerance 0.2 Da for common ions and 0.3 for cross-link ions; search in enumeration mode (exhaustive search). Search results were filtered according to the following criteria: MS1 mass tolerance window –3 to 7 ppm. Finally each MS/MS spectra was manually inspected and validated.

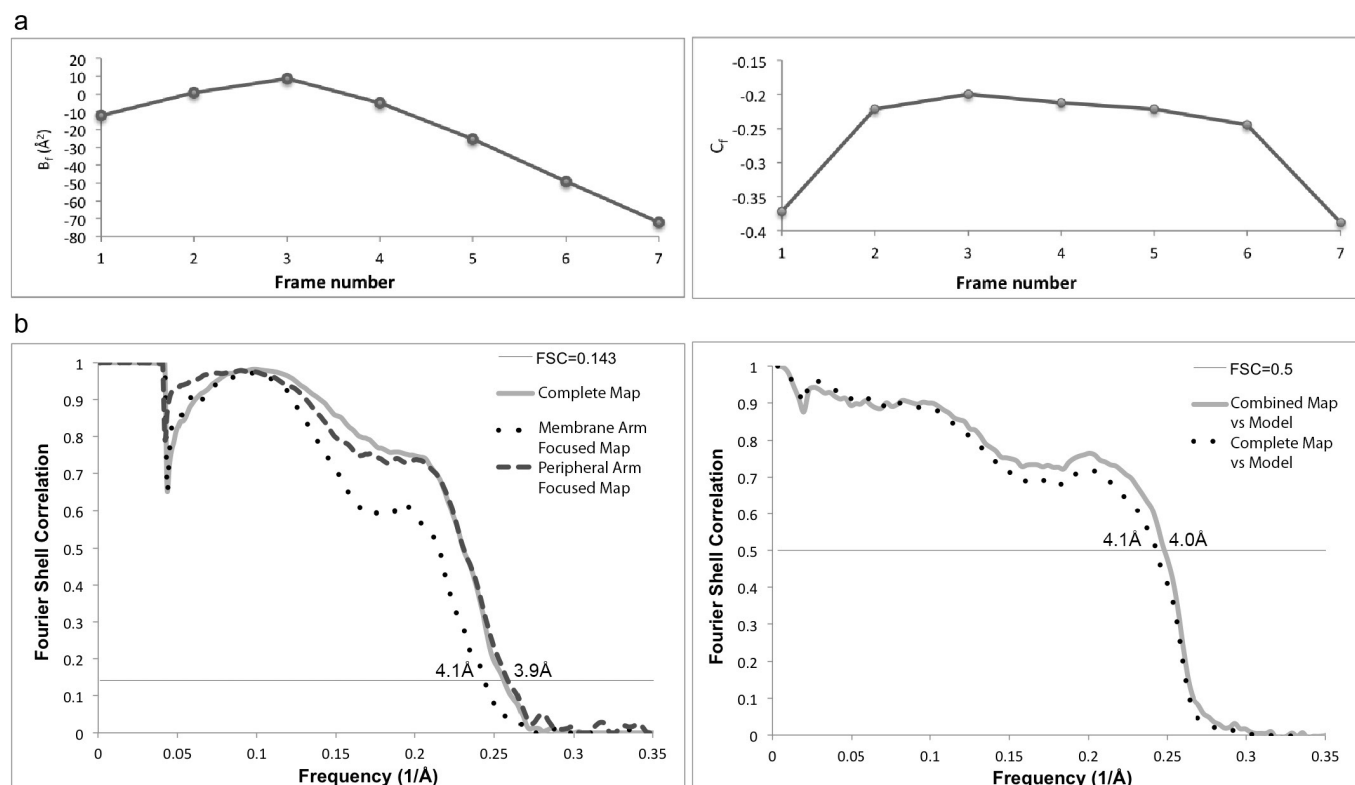
In total 218 unique cross-linked peptides were identified, of which 87 were between residues of different subunits (inter-subunit, Supplementary Table 2), 73 were between residues within the same subunit (intra-subunit, Supplementary Table 3), and 58 were clear false positives (Supplementary Table 4). False positives were

identified by comparison to all known biochemical and structural information on complex I and the cross-links that are considered false positives are either between residues that are too distant from each other (>32 Å after allowing for exposed side chain flexibility from their modelled position), located on opposite sides of the membrane or the reactive residues are buried and not solvent accessible in the intact structure. Many of the false positive cross-links are found on unstructured coils at the edges of the complex I structure indicating that they probably result from transient interactions between different complexes I during the reaction (inter-complex cross-links). True positive cross-links were more likely to be observed in more than one experiment. Some high-scoring cross-links were observed between disordered termini or loops of subunits that could not be modelled in our structure; hence the accurate determination of distance for these cross-links was not possible. Nonetheless in cases where cross-linking residues are adjacent to the modelled regions, the cross-links were considered true and are included in Extended Data Fig. 5 and Supplementary Tables 2 and 3. No cross-links were observed for any of the mitochondrially encoded core subunits, which are buried in the membrane and coated with a layer of supernumerary subunits. Good quality cross-links were observed for all supernumerary subunits except for B14.7, KFYI and AGGG. These data in conjunction with our electron microscopy maps allowed us to unambiguously assign all supernumerary subunits. Previous assignments were confirmed and importantly, subunits that previously had no known position in the complex (10 kDa, B14.5a, MWFE, B9, MNLL, SGD, ASHI, B17, AGGG and B12) have now been assigned and built (see Supplementary Discussion for more details).

32. Sazanov, L. A., Peak-Chew, S. Y., Fearnley, I. M. & Walker, J. E. Resolution of the membrane domain of bovine complex I into subcomplexes: implications for the structural organization of the enzyme. *Biochemistry* **39**, 7229–7235 (2000).
33. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
34. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
35. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
36. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
37. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
38. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
39. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
40. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* **4**, e11182 (2015).
41. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
42. Allegretti, M., Mills, D. J., McMullan, G., Kühlbrandt, W. & Vonck, J. Atomic model of the F₄₂₀-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife* **3**, e01963 (2014).
43. Meents, A., Gutmann, S., Wagner, A. & Schulze-Bries, C. Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proc. Natl Acad. Sci. USA* **107**, 1094–1099 (2010).
44. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778–795 (2009).
45. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).
46. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
47. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Res.* **32**, W321–W326 (2004).
48. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
49. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
50. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
51. Shinzawa-Itoh, K. *et al.* Bovine heart NADH-ubiquinone oxidoreductase contains one molecule of ubiquinone with ten isoprene units as one of the cofactors. *Biochemistry* **49**, 487–492 (2010).
52. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
53. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protocols* **9**, 120–137 (2014).
54. Kahraman, A., Malmström, L. & Aebersold, R. Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **27**, 2163–2164 (2011).



Extended Data Figure 1 | Image processing procedures. **a**, Representative micrograph of 2.6 k micrographs collected that all varied in defocus, ice thickness and particle count, with good quality particles circled. Scale bar, 100 nm. **b**, Representative 2D class averages obtained from reference-free classification. **c**, Classification and refinement procedures used in this study.



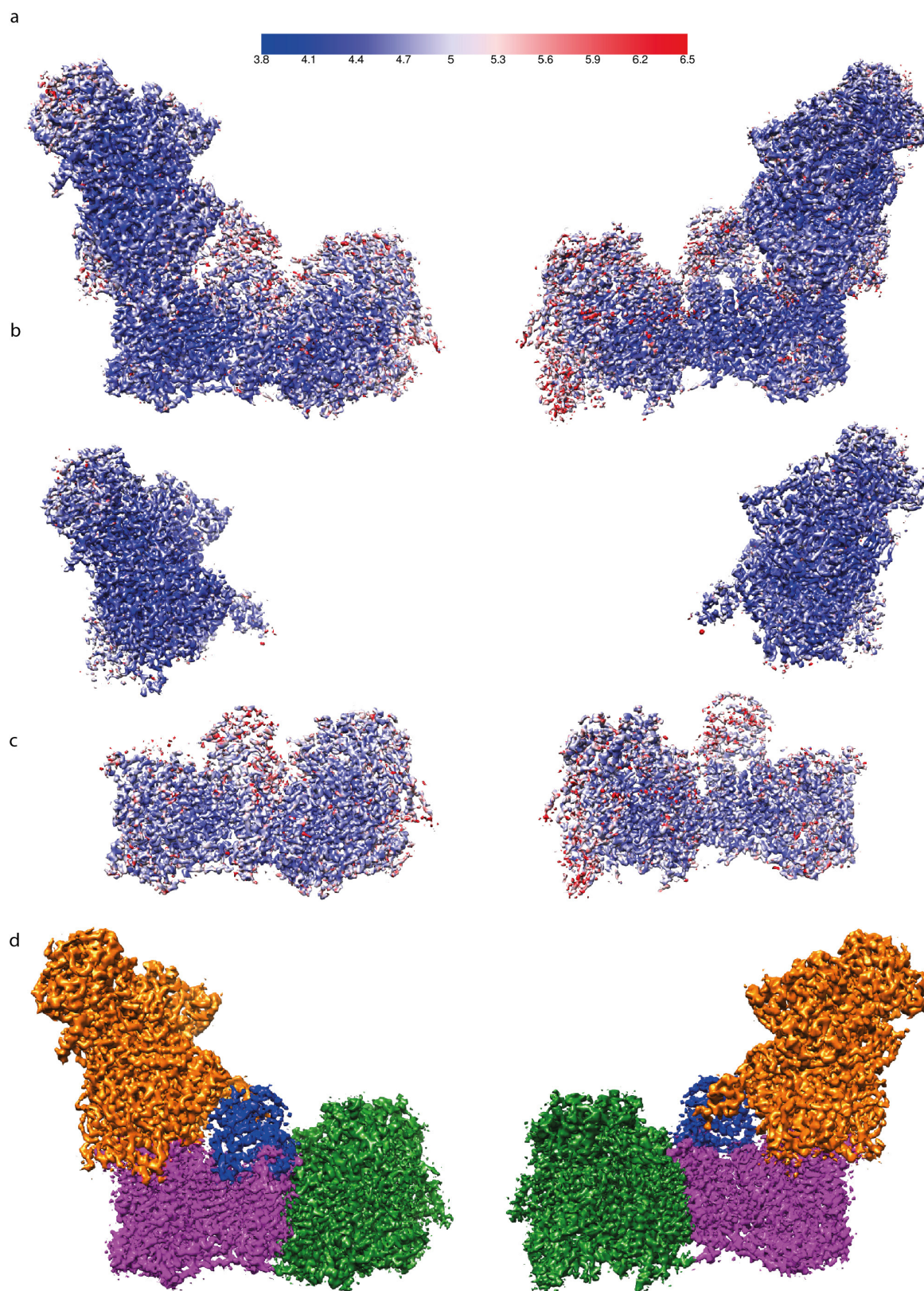
c. Statistics

Data collection	
EM	Titan Krios 300kV, FEI Falcon II
Pixel size (Å)	1.39
Defocus range (μm)	-0.5 to -3.5
Reconstruction (RELION)	
Overall	Membrane Domain
Accuracy of rotations (°)	0.573
Accuracy of translations (pixel)	0.308
B-factor from post-processing	-88
B-factor for map visualisation	-100
Final resolution (Å)	3.9
Model refinement (PHENIX)	
Resolution limit (Å)	3.9
Number of residues	8037
Map CC (whole unit cell)	0.758
Map CC (around atoms)	0.782
Rmsd (bonds)	0.009
Rmsd (angles)	1.04
Average B-factor	86.0
Validation	
All-atom clashscore	24.4
Ramachandran plot	
Outliers (%)	0.5
Allowed (%)	12.5
Favoured (%)	87.0
Rotamer outliers (%)	0.1
Molprobtity score	2.5

Extended Data Figure 2 | Image and model refinement procedures.

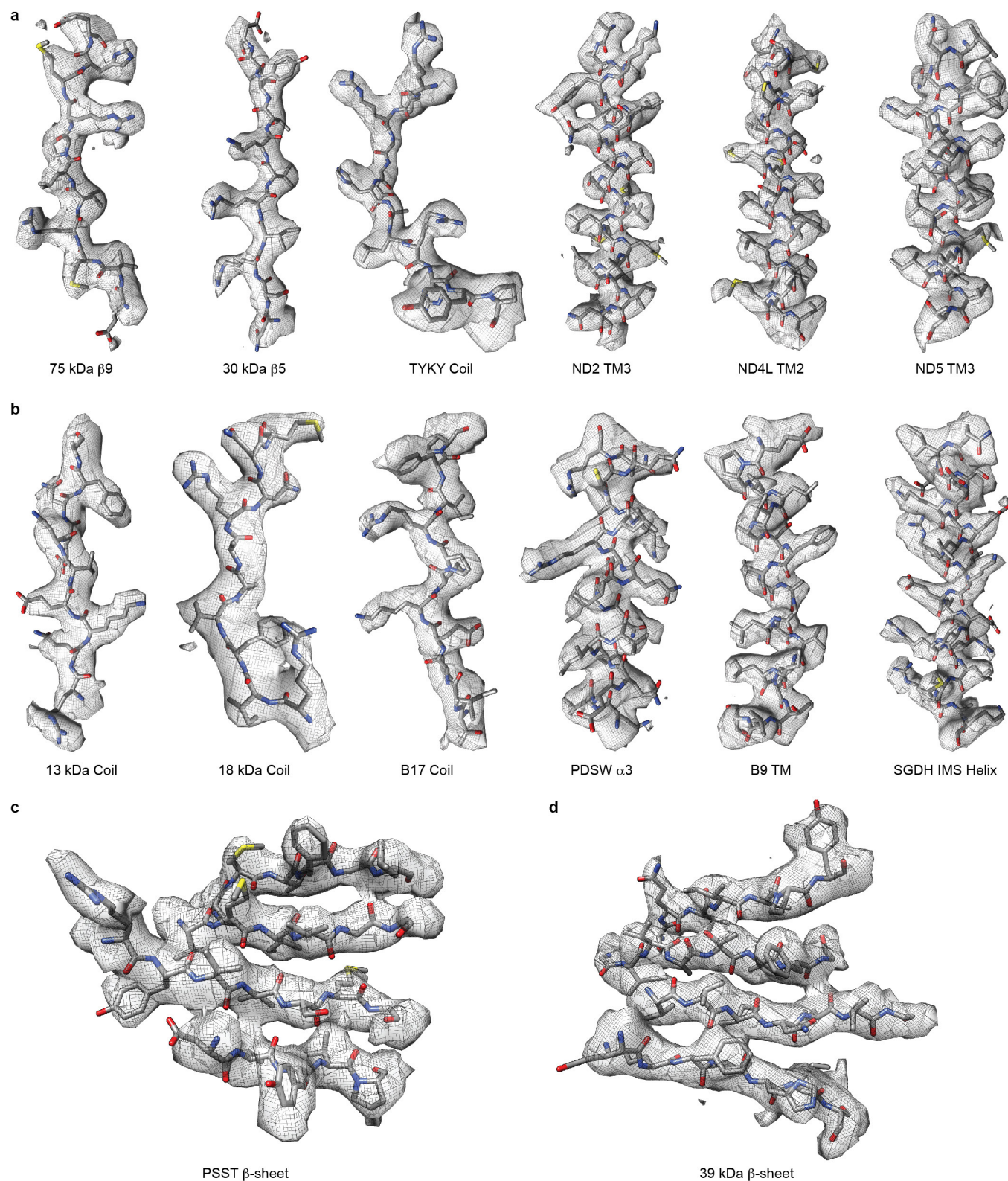
a, Radiation-damage weighting. Relative B-factors (B_r) and intercepts (C_r) from the Relion particle polishing procedure. **b**, Left, gold-standard (two halves of data refined independently) Fourier shell correlation (FSC) curves for the maps of the entire complex complete map (resolution at FSC = 0.143 is 3.9 Å), membrane-arm-focused refinement (4.1 Å

resolution) and peripheral-arm-focused refinement (3.9 Å resolution). Right, FSC curve of the combined map versus final model shows good agreement of the model with the map (FSC = 0.5 at 4.0 Å resolution). FSC curve against the entire complex complete map, which was not used in refinement, is shown as a control. **c**, Statistics of refinement.

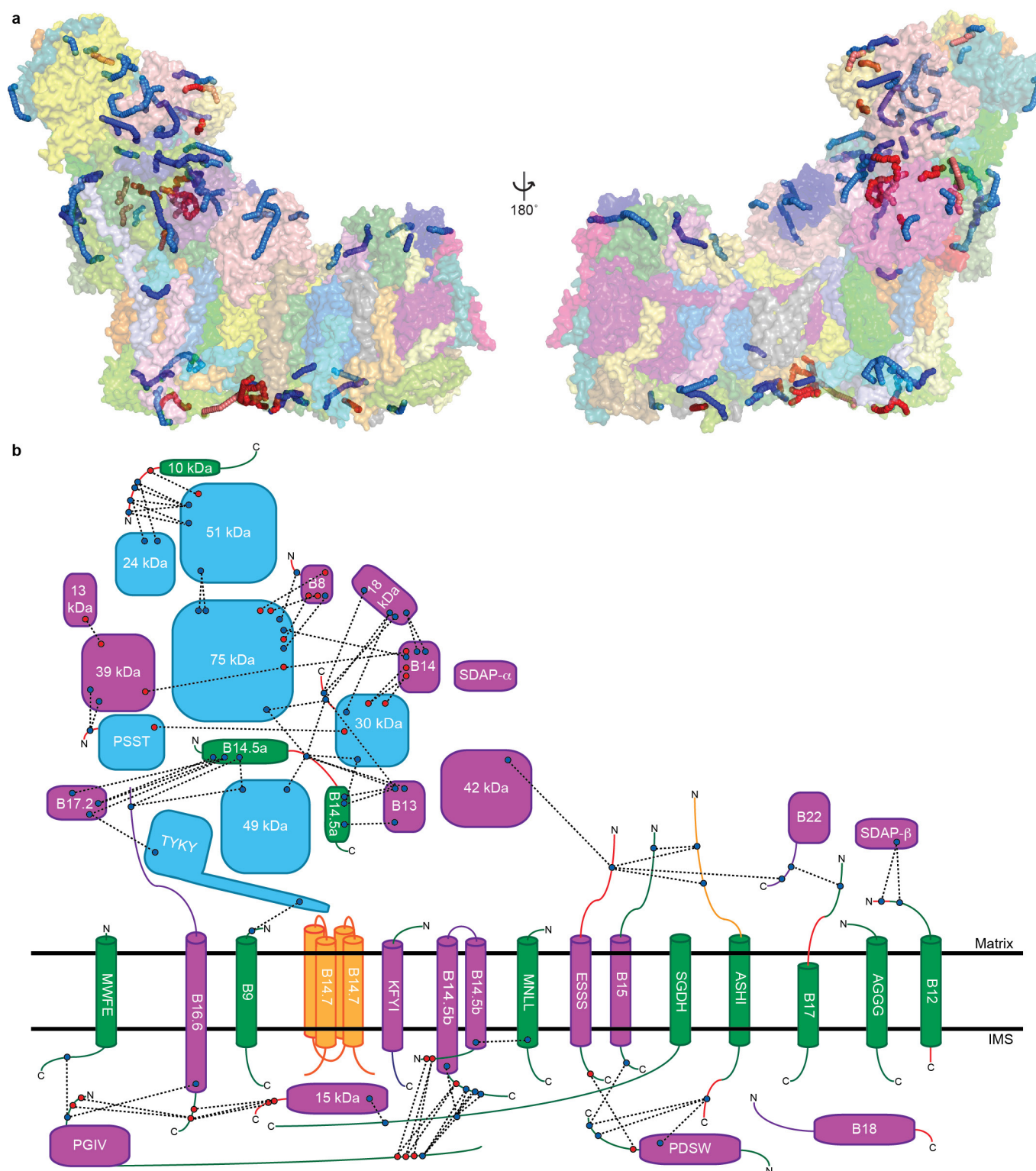


Extended Data Figure 3 | Local resolution estimation and combination of maps for model building. **a–c**, Local resolution estimation by Resmap of the entire complex I (**a**), peripheral-arm-focused refinement map (**b**) and membrane-arm-focused refinement map (**c**). Maps are coloured according to the shown resolution scale in Å. **d**, The final map was produced by combining maps with the best local resolution features; that

is, for peripheral-arm-focused refinement map (orange), for the distal part of membrane-arm-focused refinement map (green), for 42-kDa subunit map from the selected homogenous complex I class (64 k particles; blue) and the rest of the complex from the best map of the entire complex (magenta).

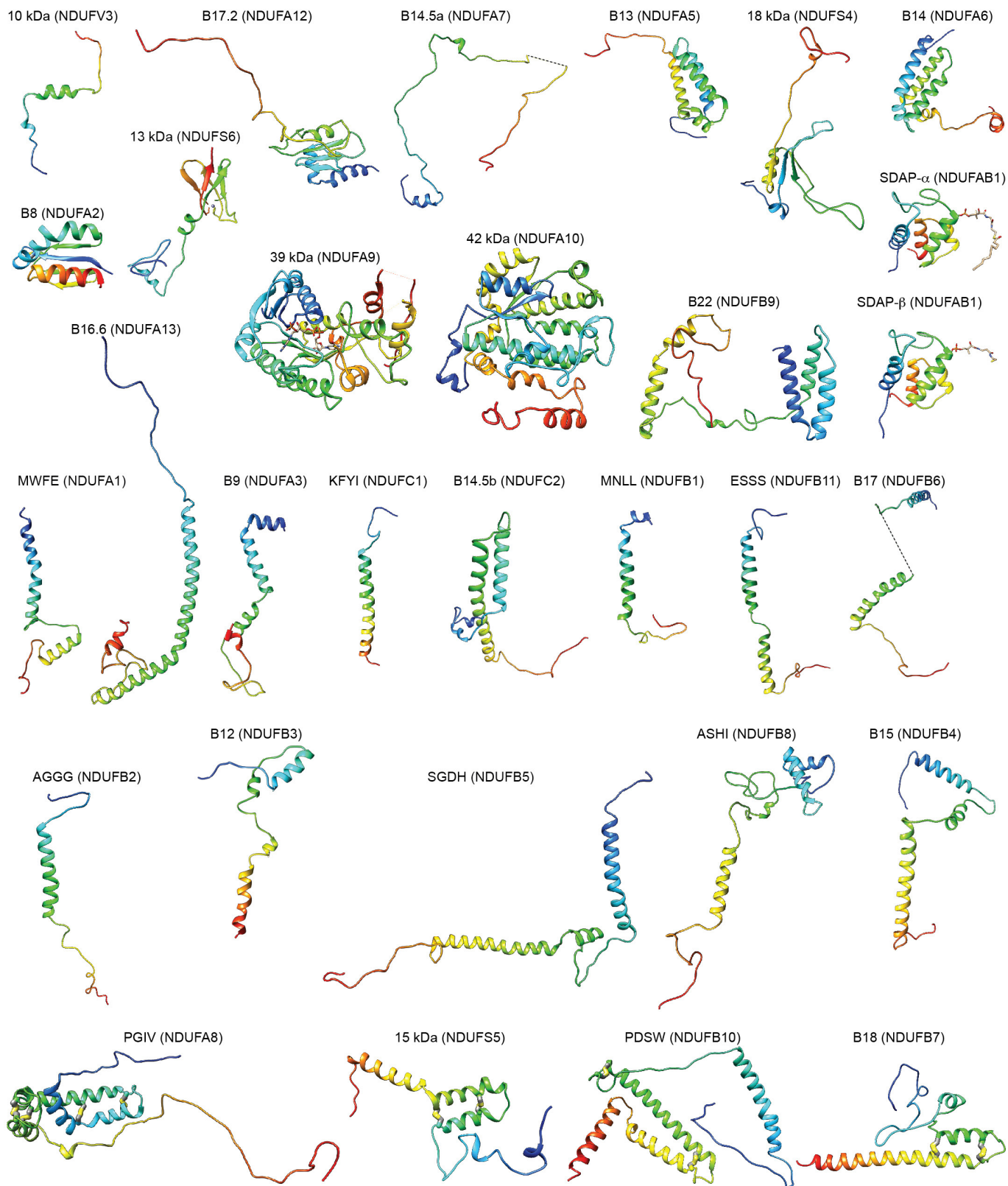


Extended Data Figure 4 | Examples of cryo-EM density. a, b, Coils and α -helices from core (a) and supernumerary (b) subunits. **c, d,** Example β -sheets from core PSST subunit (c) and supernumerary 39-kDa subunit (d). Cryo-EM density is shown with the model represented as sticks and coloured by atom with carbon in grey, oxygen in red, nitrogen in blue and sulfur in yellow.

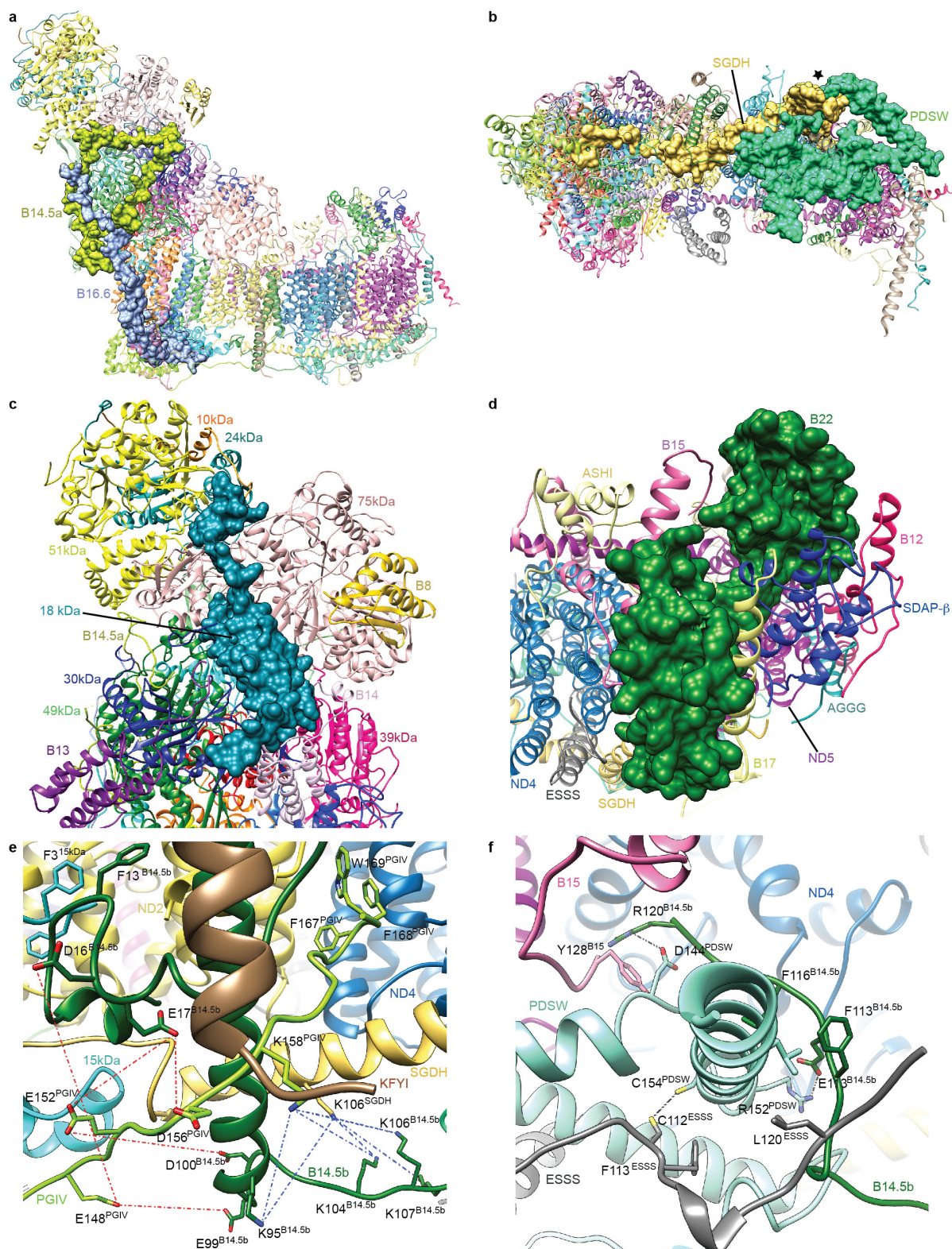


Extended Data Figure 5 | Identified cross-links. **a**, Solvent-accessible surface (SAS) representation of cross-links. Surfaces for complex I subunits are shown transparent and coloured as in Fig. 1. Shortest SAS paths calculated using Xwalk⁵⁴ are shown for cross-links as coloured worms with inter-subunit lysine reactive cross-links in blue, inter-subunit acid reactive cross-links in red, intra-subunit lysine reactive cross-links in light blue, intra-subunit acid reactive cross-links in light red. **b**, Inter-subunit cross-link schematic. Complex I subunits are shown in a similar orientation as in **a**. Left panel with core subunits cyan, previously assigned supernumerary subunits in magenta, newly assigned or newly built regions

of supernumerary subunits in green, poly-alanine regions in orange and unmodelled regions in red. Observed cross-links are indicated by dashed black lines between either blue circles (lysine reactive cross-links) or red circles (acid reactive cross-links). No cross-links were observed to the core subunits of the membrane arm and hence they were omitted for clarity. The horizontal black lines indicate the approximate boundaries of the inner mitochondrial membrane. Subunits B14.7, B15 and ASH1 are shown as being behind the membrane boundaries as they are found on the opposite (far) side of the membrane arm.

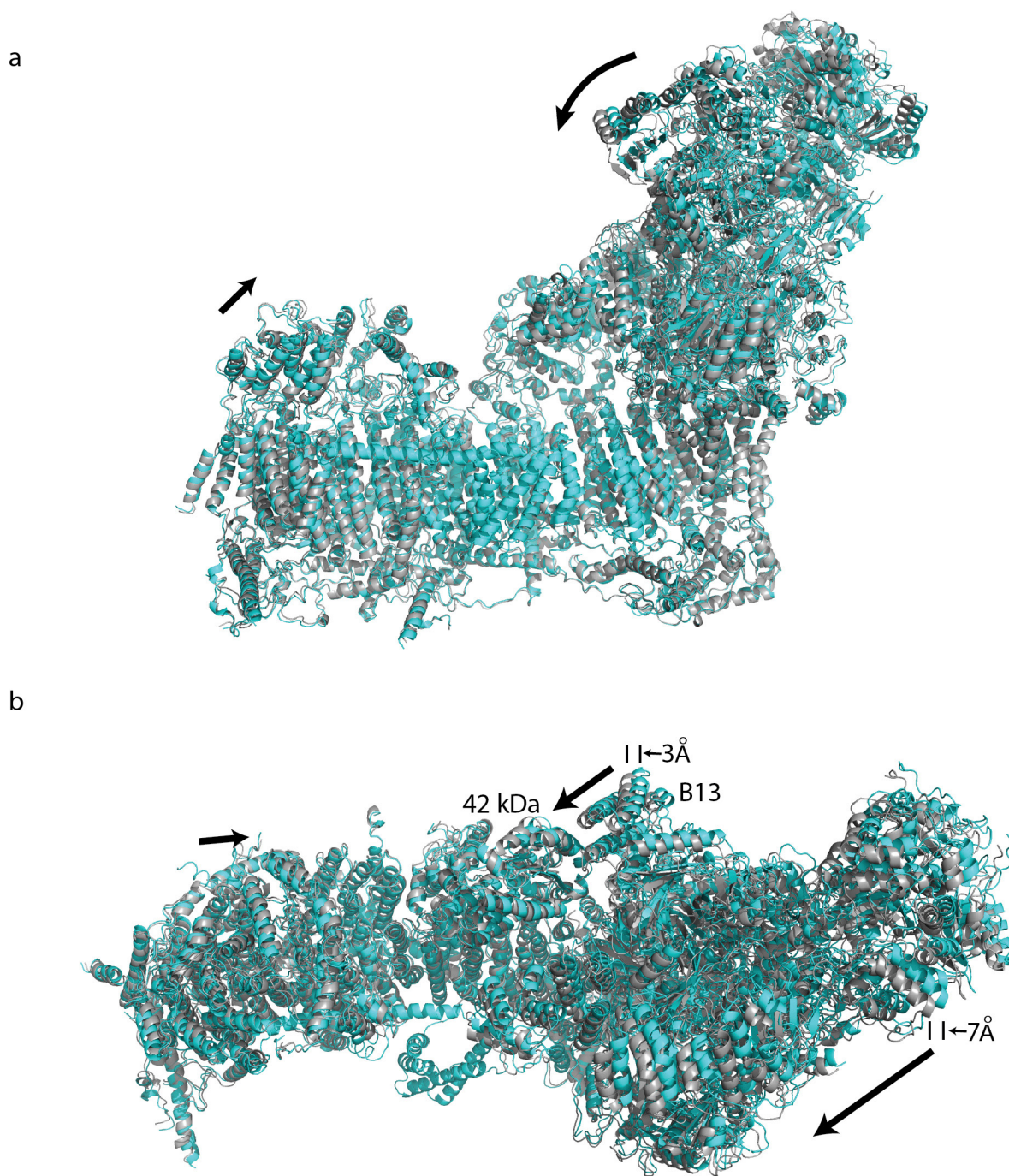


Extended Data Figure 6 | Folds of supernumerary subunits. Subunits are shown in cartoon representation, coloured blue to red from N to C terminus. Disulfide bridges are shown as sticks with sulfur in yellow.



Extended Data Figure 7 | Examples of supernumerary subunits interactions. **a**, Side view of complex I showing surfaces for subunits B14.5a and B16.6. **b**, IMS view of complex I showing surfaces for subunits SGDH and PDSW. The point at which the two subunits are intertwined is marked with a star. **c**, View of the hydrophilic arm looking from above the membrane arm. The surface of the 18-kDa subunit that spans the hydrophilic arm is shown. **d**, Matrix view of the tip of the membrane arm with the surface of supernumerary subunit B22 shown. **e**, Close up of the centre of the membrane arm on the IMS side. This region contains many interactions between supernumerary subunits and the side chains of

residues involved are shown. The region is also a hot spot for cross-links, the side chains involved are shown and cross-links are indicated with dashed lines (acid cross-links: red; basic cross-links: blue). **f**, Close up of the C-terminal helix of supernumerary subunit PDSW at the centre of the membrane arm on the IMS side. This helix extends away from complex I and is encircled by the C termini of supernumerary subunits B14.5b, ESSS and B15. The side chains of residues involved in stabilizing interactions are shown. A possible disulfide bond between PDSW (Cys154) and ESSS (Cys112) and stabilizing salt bridges are indicated by dashed lines. Subunits are coloured as in Fig. 1.



Extended Data Figure 8 | Comparison of ‘open’ and ‘closed’ 3D class structures. **a, b,** Side (**a**) and top (**b**) view from the matrix for the alignment of the open class structure (in cyan) and closed class structure (in grey). To generate the closed class structure, the final structure of the open class was refined in real space in Phenix (5 macro cycles with morphing at each cycle) against 4.6 Å map of the closed class (Extended Data Fig. 1). All of the α -helices were well fit into density, but owing to low resolution of the closed class no further refinement was performed and the comparison of structures involves only the relative positions of secondary

structure elements. The two structures were aligned via transmembrane core subunits and are displayed as cartoon models. In the closed class the peripheral arm undergoes a hinge-like motion around the Q site towards the tip of the membrane arm, with the direction of shift indicated by the arrow in **b**. As a result, subunit B13 moves ~ 3 Å closer to the 42-kDa subunit, allowing for direct contacts. The shift is larger at the periphery, reaching 7 Å at the tip of the peripheral arm. Additionally, subunit ND5 and its matrix bulge move about 3 Å towards peripheral arm.

Extended Data Table 1 | Summary of the model

	Subunit Name Bovine / Human	Chain	Total residues / range built	Poly-ALA model	Un-modelled residues	% atomic model	TMHs	Cofactors / Notes
C o r e	51kDa / NDUFV1	1	445 / 7-438	-	1-6, 439-445	97.1	-	FMN N3 (4Fe[51])
	24kDa / NDUFV2	2	217 / 3-216	-	1-2, 217	98.6	-	N1a (2Fe[24])
	75kDa / NDUFS1	3	704 / 6-693	-	1-5, 694-704	97.7	-	N1b (2Fe[75]) N4 (4Fe[75]C) N5 (4Fe[75]H)
	49kDa / NDUFS2	4	430 / 44-430	-	1-43	90.0	-	-
	30kDa / NDUFS3	5	228 / 7-214	-	1-6, 215-228	91.2	-	-
	PSST / NDUFS7	6	179 / 25-179	-	1-24	86.6	-	N2 (4Fe[PS])
	TYKY / NDUFS8	9	176 / 1-176	-	-	100	-	N6a (4Fe[TY]1) N6b (4Fe[TY]2)
	ND1	H	318 / 1-318	-	-	100	8	-
	ND2	N	347 / 1-347	-	-	100	11	(antiporter-like)
	ND3	A	115 / 1-115	-	-	100	3	-
	ND4	M	459 / 1-459	-	-	100	14	(antiporter-like)
	ND4L	K	98 / 1-86	-	87-98	87.8	3	-
	ND5	L	606 / 1-568	509-598	599-606	84.0	16	(antiporter-like)
S u p e r n u m e r a r y	ND6	J	175 / 1-108, 123-175	77-122	-	73.7	5	-
	10kDa / NDUFV3	a	75 / 34-74	-	1-33, 75	54.7	-	-
	13kDa / NDUFS6	b	96 / 1-95	-	96	99.0	-	Zn ²⁺
	18kDa / NDUFS4	c	133 / 11-133	-	1-10	92.5	-	-
	39kDa / NDUF A9	d	345 / 1-252, 277-338	325-338	253-276, 339-345	87.0	-	NADPH
	B8 / NDUF A2	e	98 / 13-96	-	1-12, 97-98	85.7	-	Thioredoxin fold
	B13 / NDUF A5	f	115 / 4-115	-	1-3	97.4	-	-
	B14 / NDUF A6	g	127 / 14-127	-	1-13	89.8	-	LYR protein
	B14.5a / NDUF A7	h	112 / 1-71, 89- 112	-	72-88	84.8	-	-
	B17.2 / NDUF A12	i	145 / 1-144	-	145	99.3	-	-
	SDAP- α / NDUFAB1	j	88 / 3-87	-	1-2, 88	96.6	-	Acyl carrier protein phosphopantetheine
	42kDa / NDUF A10	k	320 / 1-320	227-320	-	70.9	-	Nucleoside kinase family Quadruple CX ₉ C 2 CHCH domains
	15kDa / NDUFS5	l	105 / 1-95	-	96-105	90.5	-	-
	B9 / NDUF A3	m	83 / 4-83	-	1-3	96.4	1	STMD
	B12 / NDUF B3	n	97 / 13-86	83-86	1-12, 87-97	73.2	1	STMD
	B14.5b / NDUF C2	o	120 / 1-120	-	-	100	2	-
	B15 / NDUF B4	p	128 / 17-73, 95-128	17-27, 74-94	1-16	62.5	1	STMD
	B16.6 / NDUF A13	q	143 / 4-143	-	1-3	97.9	1	Identical to GRIM-19, STMD
	B17 / NDUF B6	r	127 / 1-37, 63- 118	63-71	38-62, 119-127	66.1	1	STMD
	B18 / NDUF B7	s	136 / 5-122	5-15	1-4, 123-136	78.7	-	Double CX ₉ C CHCH domain
	B22 / NDUF B9	t	178 / 9-174	-	1-8, 175-178	93.3	-	LYR protein
	AGGG / NDUF B2	u	72 / 3-68	-	1-2, 69-72	91.7	1	STMD
	ASHI / NDUF B8	v	158 / 1-84, 101-143	1-100	144-158	27.2	1	STMD
	ESSS / NDUF B11	w	125 / 38-123	-	1-37, 124-125	68.8	1	STMD
	KFYI / NDUF C1	x	49 / 1-48	-	49	98.0	1	STMD
	MNLL / NDUF B1	y	57 / 4-56	-	1-3, 57	93.0	1	STMD
	MWFE / NDUF A1	z	70 / 2-70	-	1	98.6	1	STMD
	PDSW / NDUF B10	Z	175 / 3-173	-	1-2, 174-175	97.7	-	-
	PGIV / NDUF A8	Y	171 / 1-171	-	-	100	-	Quadruple CX ₉ C 2 CHCH domains
	SDAP- β / NDUFAB1	X	88 / 1-88	-	-	100	-	Acyl carrier protein phosphopantetheine
	SGDH / NDUF B5	W	143 / 5-143	-	1-4	97.2	1	STMD
	B14.7 / NDUF A11	V	140 / 7-125	7-125	21	0	4	-
	Total		8516 / 8037	518	479	88.3	78	

Extended Data Table 2 | Summary of interactions between subunits of ovine complex I

Subunit	Surface area, Å ²	Buried area, Å ²	$\Delta G_{\text{int}}^{\text{int}}$ kcal/mol	N _{HB}	N _{SB}	Interacting subunits (Descending buried area order, core subunits in bold)
Entire complex summary						
	297384.9	259150.4	-1838.5			
Core subunits						
51kDa	19017.7	6988.3	-12.38	77	18	24kDa, 75kDa , 10kDa, 18kDa, B14.5a
24kDa	14271.6	5144.8	-34.27	43	15	51kDa , 10kDa, 75kDa , 18kDa, 13kDa
75kDa	29733.9	10317.9	-23.97	102	43	18kDa, 51kDa , B8, 49kDa , 24kDa , 30kDa , B17.2, B14, TYKY , 13kDa, B14.5a, 39kDa, 10kDa
49kDa	17459.5	11703.0	-39.24	93	37	30kDa , TYKY , PSST , B14.5a, ND1 , B16.6, 75kDa , B13, ND3 , 13kDa, B14
30kDa	14832.7	9972.2	-40.37	79	34	49kDa , B13, 18kDa, B14.5a B14 PSST , 75kDa , 39kDa, TYKY
PSST	9483.5	6258.3	-26.86	47	31	49kDa , ND1 , TYKY , 30kDa , 39kDa, ND3 , B17.2, B14
TYKY	13499.4	10446.6	-47.00	87	33	49kDa , B17.2, PSST , B14.5a, 13kDa, ND1 , 75kDa , B16.6, B9, 30kDa , 18kDa, 39kDa, MWFE
ND1	16976.8	11161.2	-98.77	61	21	ND3 , MWFE, PSST , 49kDa , ND6 , B16.6, TYKY , B9, PGIV, B17.2
ND2	16531.5	9336.0	-106.34	19	1	ND4L , 42kDa, ND4 , B14.5b, 15kDa, ND5 , SGDHD
ND3	10944.8	6994.6	-81.41	15	0	ND1 , ND6 , ND4L , PSST , 49kDa, B9, 39kDa, B14, B16.6
ND4	18758.1	11100.2	-98.26	42	3	ND5 , ESSS, ND2 , B15, SGDHD, MNLL, PDSW, B22, PGIV, B14.5b, ASHI
ND4L	6435.4	4638.4	-64.53	7	2	ND6 , ND2 , 15kDa, ND3
ND5	26832.2	10837.8	-94.45	47	5	ND4 , AGGG, B22, PDSW, ASHI, B17, B18, B12, B15, ND2 , SDAP- β , SGDHD, B14.7
ND6	12734.4	8272.0	-95.71	19	2	ND4L , ND3 , ND1 , 15kDa, B16.6, MWFE
Supernumerary subunits						
10kDa	4892.2	2471.0	-15.11	23	9	51kDa , 24kDa , 18kDa, 75kDa
13kDa	7651.1	2991.2	-10.11	16	6	TYKY , B17.2, 39kDa, 75kDa , 49kDa , 24kDa
18kDa	11083.8	5848.7	-11.52	57	20	75kDa , 30kDa , B14, 51kDa, 39kDa, 10kDa, TYKY , 24kDa , B17.2
39kDa	16659.0	3814.3	-5.70	20	11	PSST , 13kDa, 30kDa , B14, 18kDa, ND3 , 75kDa , TYKY
B8	5628.2	1196.4	-2.44	13	10	75kDa
B13	8653.4	2370.0	-15.60	14	5	30kDa , 49kDa , B14.5a
B14	9627.2	4586.9	-9.29	31	19	SDAP- α , 30kDa , 18kDa, 75kDa , 39kDa, ND3 , 49kDa , PSST
B14.5a	11473.6	5753.0	-40.33	49	17	49kDa , 30kDa , TYKY , 75kDa , B17.2, B13, B16.6, 51kDa
B17.2	12313.0	4593.5	-25.36	40	9	TYKY , 13kDa, 75kDa , B14.5a, PSST , ND1 , 18kDa
SDAP- α	5667.7	731.3	-3.93	7	12	B14
SDAP- β	6074.4	2554.8	-9.23	22	26	B22, B17, B12, ND5 , AGGG
42kDa	16761.9	1942.1	-19.89	2	0	ND2 , KFYI, B14.5b
15kDa	9228.8	5236.4	-29.49	27	9	SGDHD, B16.6, ND2 , ND6 , ND4L , PGIV, B14.5b
B9	8142.1	3420.4	-27.55	19	8	PGIV, ND1 , B16.6, TYKY , ND3 , SGDHD
B12	7188.3	2427.6	-13.38	14	7	ND5 , AGGG, SDAP- β , B22
B14.5b	10935.4	5873.6	-39.62	31	8	ND2 , KFYI, PDSW, SGDHD, PGIV, ND4 , 15kDa, ESSS, 42kDa, B15
B15	10929.5	3772.9	-15.63	16	0	ND4 , ASHI, B22, ND5 , PDSW, B14.5b
B16.6	14451.0	7839.8	-51.27	42	12	PGIV, 15kDa, ND1 , 49kDa , ND6 , MWFE, B9, TYKY , B14.5a, ND3 , SGDHD
B17	10137.0	4528.8	-24.39	29	6	B22, PDSW, ND5 , B18, SDAP- β , SGDHD
B18	9860.6	3512.3	-6.33	24	6	ND5 , AGGG, ASHI, B17, PDSW
B22	14868.0	5662.9	-32.67	25	15	ND5 , B17, SDAP- β , B15, SGDHD, B12, ND4 , ASHI
AGGG	7273.2	3125.3	-26.86	17	7	ND5 , B18, B12, SDAP- β , ASHI
ASHI	11797.4	2971.5	-22.74	10	1	B15, ND5 , B18, ND4 , AGGG, B22
ESSS	8835.5	4764.1	-39.36	21	5	ND4 , PDSW, SGDHD, B14.5b
KFYI	5365.5	1349.6	-12.98	0	2	B14.5b, 42kDa
MNLL	5714.8	1719.4	-14.42	9	5	ND4 , SGDHD, PDSW
MWFE	6638.2	3788.4	-29.68	29	4	ND1 , PGIV, B16.6, ND6 , TYKY
PDSW	15770.6	7444.1	-35.07	54	8	ESSS, SGDHD, ND5 , B17, B14.5b, ND4 , B18, B15, MNLL
PGIV	14545.4	6552.6	-38.16	49	28	B16.6, B9, MWFE, SGDHD, B14.5b, 15kDa, ND4 , ND1
SGDHD	14101.4	8031.0	-55.12	40	10	ND4 , PDSW, 15kDa, ESSS, B14.5b, MNLL, PGIV, B22, ND2 , B17, ND5 , B16.6, B9

Analysis was performed using the PISA server (<http://www.ebi.ac.uk/pdbe/pisa/>). ΔG_{int} indicates the solvation free energy gain upon formation of the assembly. N_{HB}, number of hydrogen bonds at the interface; N_{SB}, number of salt bridges at the interface.

X-ray structure of the human $\alpha 4\beta 2$ nicotinic receptor

Claudio L. Morales-Perez¹, Colleen M. Noviello¹ & Ryan E. Hibbs¹

Nicotinic acetylcholine receptors are ligand-gated ion channels that mediate fast chemical neurotransmission at the neuromuscular junction and have diverse signalling roles in the central nervous system. The nicotinic receptor has been a model system for cell-surface receptors, and specifically for ligand-gated ion channels, for well over a century^{1,2}. In addition to the receptors' prominent roles in the development of the fields of pharmacology and neurobiology, nicotinic receptors are important therapeutic targets for neuromuscular disease, addiction, epilepsy and for neuromuscular blocking agents used during surgery^{2–4}. The overall architecture of the receptor was described in landmark studies of the nicotinic receptor isolated from the electric organ of *Torpedo marmorata*⁵. Structures of a soluble ligand-binding domain have provided atomic-scale insights into receptor–ligand interactions⁶, while high-resolution structures of other members of the pentameric receptor superfamily provide touchstones for an emerging allosteric gating mechanism⁷. All available high-resolution structures are of homopentameric receptors. However, the vast majority of pentameric receptors (called Cys-loop receptors in eukaryotes) present physiologically are heteromeric. Here we present the X-ray crystallographic structure of the human $\alpha 4\beta 2$ nicotinic receptor, the most abundant nicotinic subtype in the brain. This structure provides insights into the architectural principles governing ligand recognition, heteromer assembly, ion permeation and desensitization in this prototypical receptor class.

The $\alpha 4\beta 2$ receptor is known to assemble in two functional subunit stoichiometries, $3\alpha:2\beta$ and $2\alpha:3\beta$. The latter stoichiometry has a ~ 100 -fold higher affinity for both acetylcholine and nicotine,

lower single channel conductance and calcium permeability, and its expression is selectively upregulated by nicotine^{8–10}. We used a small-scale fluorescence-based approach to optimize conditions for protein expression and purification that would yield the $2\alpha:3\beta$ form¹¹. Growth of well-diffracting crystals required deleting most of the intracellular domain between transmembrane spans M3 and M4 in both subunits (Extended Data Figs 1 and 2). This crystallized receptor construct, referred to here as $\alpha 4\beta 2$, retains function comparable to the full-length protein, as discussed later. The best diffracting crystals were obtained by co-crystallization with nicotine and a cholesterol analogue, and allowed for collection of a complete data set to 3.9 Å resolution (see Methods and Extended Data Table 1).

The structure of the $\alpha 4\beta 2$ receptor was solved by molecular replacement (see Methods). Subunit identities were initially assigned based on features in electron density maps from the vicinity of the neurotransmitter-binding pocket (Extended Data Fig. 3a, b). To interrogate subunit identity further, we co-crystallized the receptor with 5-Iodo-A-85380, a potent agonist that, like acetylcholine and nicotine, is expected to bind only at α – β interfaces¹². From a low-resolution isomorphous data set we observed iodine anomalous signal in only the two assigned α – β interfaces (Extended Data Fig. 3c). After finalizing subunit assignment, electron density maps were of sufficient quality to build and refine nearly all of the extracellular and transmembrane domains, as well as a portion of the intracellular domain (Extended Data Figs 1 and 3).

The $\alpha 4\beta 2$ receptor resembles a cylinder formed from five subunits in a pseudo-symmetrical arrangement about the channel axis. The crystal structure reveals a subunit ordering of α – β – β – α – β around the pentameric ring (Figs 1a, b), consistent with functional studies of

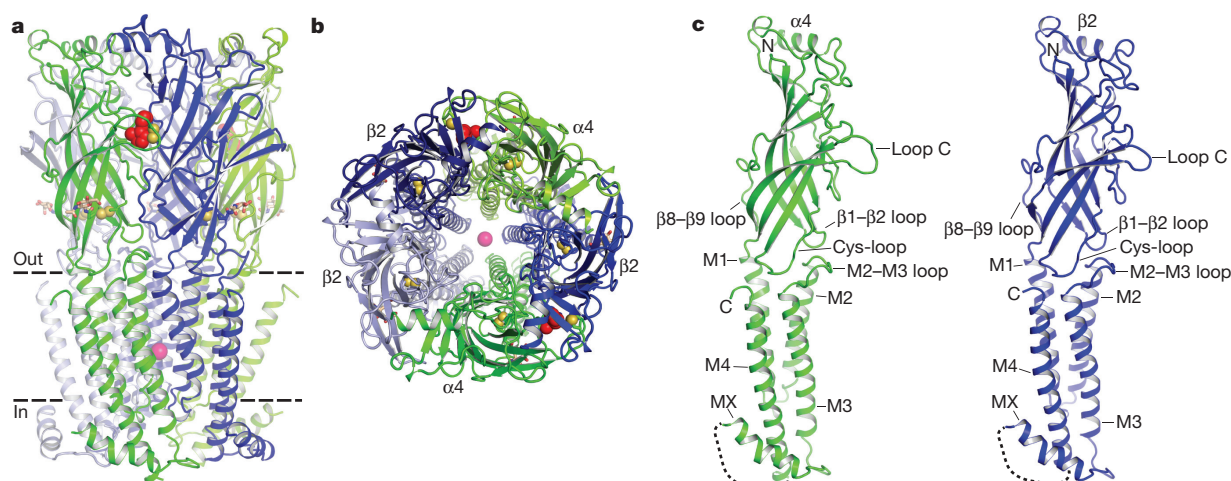


Figure 1 | Architecture of the $\alpha 4\beta 2$ nicotinic receptor. **a**, View parallel to the plasma membrane. $\alpha 4$ subunits are in green and $\beta 2$ in blue. Nicotine (red) and sodium (pink) are represented as spheres. The Cys-loop and loop C disulfide bonds are shown as yellow spheres. N-linked glycans (brown) are shown as sticks. Dashed lines indicate approximate membrane

position. **b**, View perpendicular to the plasma membrane looking from the extracellular side. **c**, Orientation as in **a** of the individual subunits. Unmodelled residues from the intracellular domain are represented as a dashed line.

¹Departments of Neuroscience and Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

concatameric receptors¹³. The $\alpha 4$ and $\beta 2$ subunits share 59% amino acid sequence identity and adopt similar backbone conformations (Fig. 1c and Extended Data Fig. 4a, b). Each subunit comprises a large extracellular domain with an N-terminal α -helix and ten β -strands that wrap inwards to form a sandwich. The C-terminal bundle comprises three transmembrane α -helices (M1–M3), an amphipathic or intracellular MX helix, and a final transmembrane α -helix (M4). The overall architecture is similar to that found in the other Cys-loop receptor family members of known structure (Extended Data Fig. 4c and Extended Data Table 2)⁷. The MX helix, about which comparatively little structural information is available, closely resembles the conformation observed in the 5-HT₃ receptor (5-HT₃R) structure (Extended Data Fig. 4c)¹⁴. The Cys-loop receptor superfamily takes its name from a conserved disulfide bond linking the $\beta 6$ and $\beta 7$ strands in the extracellular domain. A second disulfide bond is formed between adjacent cysteines at the tip of loop C in the $\alpha 4$ subunits (Extended Data Figs 3a, b and 5g–i), a feature that defines nicotinic receptor α subunits and is absent in all other Cys-loop receptors¹⁵. Electron density was observed for nicotine at the two α – β interfaces in the extracellular domain and for a single N-acetylglucosamine residue linked to a conserved asparagine in the Cys-loop of each subunit (Extended Data Fig. 3f, g). The interior surface of the receptor begins at a large extracellular vestibule that narrows into a funnel-shaped transmembrane channel defined by the pore-lining M2 α -helices; mutations in this region

are linked to autosomal-dominant nocturnal frontal lobe epilepsy (Extended Data Fig. 1)⁴. A strong electron density peak in the pore was modelled speculatively as a combination of Na⁺ ion and water in an arrangement similar to that seen in a prokaryotic pentameric receptor, GLIC¹⁶ (see Methods and Extended Data Fig. 3h, i). The channel is in a desensitized, non-conducting conformation most similar to that observed in the GABA_AR structure¹⁷; however, the overall receptor conformation is distinct.

Nicotine activity in the brain, including its reinforcing properties that lead to addiction, is mediated principally by $\alpha 4\beta 2$ receptors^{18,19}. To validate the receptor constructs used in crystallization, we quantified the binding affinities of a panel of ligands for the purified receptor (Fig. 2a and Extended Data Fig. 2d). Among the three classes of subunit interfaces, we observed electron density for nicotine only at the α – β interfaces (Fig. 2b). The ligand was positioned based on the strong omit electron density (6.8–8.0 σ ; Extended Data Fig. 3f, g) and comparison with the high-resolution structure of the acetylcholine-binding protein (AChBP) in complex with nicotine (Extended Data Fig. 6)⁶. We first analysed interactions of nicotine with the receptor and then compared the positions of corresponding residues at non- α – β interfaces to understand principles of binding selectivity.

Nicotine binds in the classical neurotransmitter site at the α – β interface, almost fully buried from solvent. The $\alpha 4$ subunit forms the (+) side of the binding pocket and the $\beta 2$ subunit forms the (–) side (Fig. 2b, c). Three loops from each side of the interface contribute to

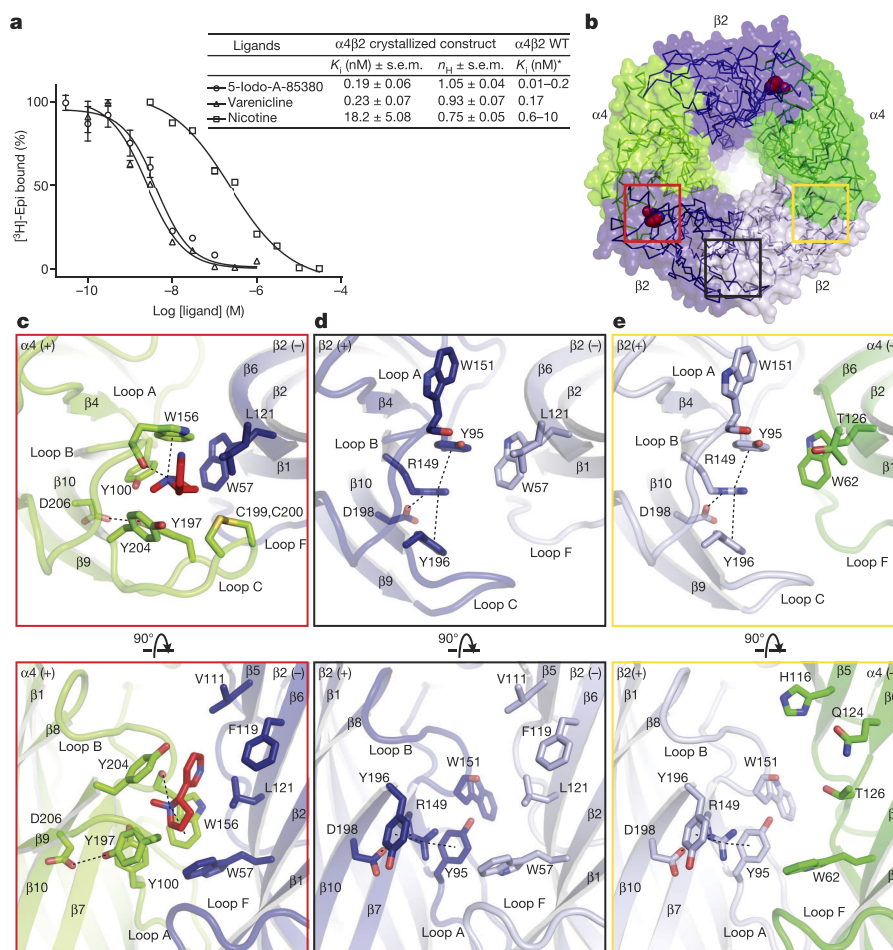


Figure 2 | Neurotransmitter-binding site. **a**, Competition experiments against [³H]-epibatidine. Calculated inhibition constant (K_i) values assume a K_d for [³H]-epibatidine of 96 pM (Extended Data Fig. 2d). $n = 4$ independent experiments. n_H , Hill coefficient. Error bars are standard error of the mean (s.e.m.). ^{*} K_i indicates published range of the ligands against wild-type (WT) $\alpha 4\beta 2$. **b**, Extracellular view, with coloured boxes

indicating the three different interface classes. **c–e**, Architectural details of interfaces boxed in **b**. The top row is from the same orientation as in **b**. Nicotine and interacting residues are shown as sticks. Potential hydrogen bonding and cation– π interactions are represented as dashed lines (2.7–5 Å). In the bottom row, the loop C backbone is hidden to aid in clarity.

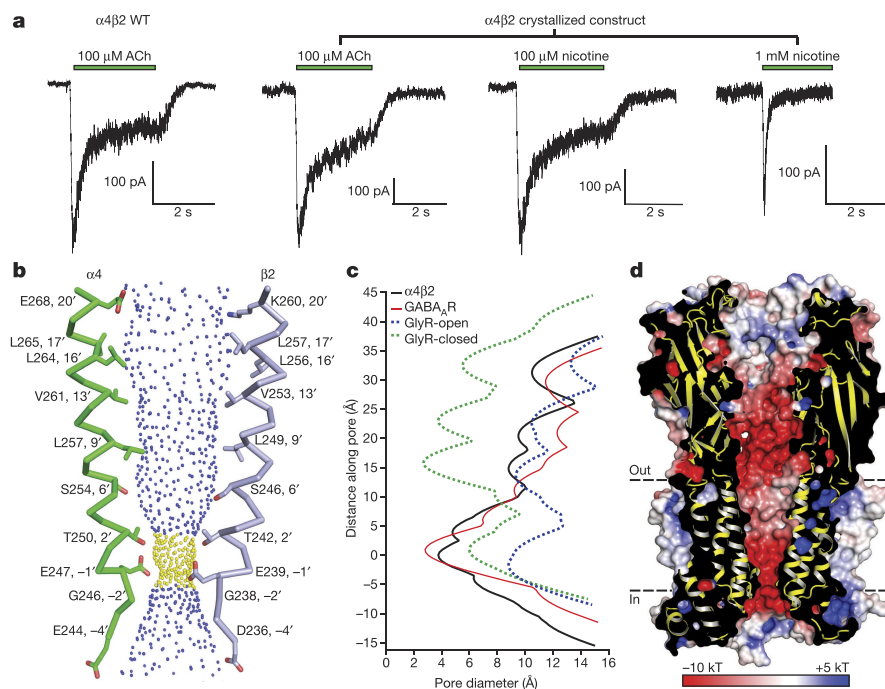


Figure 3 | Ion permeation pathway. **a**, Patch-clamp recordings of the wild type (WT) and crystallized $\alpha 4\beta 2$ receptor. ACh, acetylcholine. **b**, M2 α -helices from opposing $\alpha 4$ and $\beta 2$ subunits with side chains shown for pore-lining residues. Blue spheres indicate pore diameters >5.6 Å; yellow are >2.8 Å and <5.6 Å. **c**, Pore diameter for the $\alpha 4\beta 2$ receptor and representative Cys-loop receptors in distinct functional states: desensitized/closed (GABA_AR plus benzamidine; Protein Data

Bank (PDB) accession 4COF), activated/open (GlyR plus glycine; PDB accession 3JAE) and resting/closed (GlyR plus strychnine; PDB accession 3JAD). Structures were aligned using the M2 helix 9' leucine, which occurs at $y \approx 15$ Å. The zero value along the y axis in the plot is aligned with the α -carbon of the M2 helix -1' glutamate residue in $\alpha 4\beta 2$. **d**, Cutaway of the receptor showing the permeation pathway coloured by electrostatic potential.

binding of orthosteric ligands, A, B and C from the (+) side, and D, E and F from the (-) side. Residues from loops A–E form a tightly packed aromatic box surrounding nicotine, with the floor formed by Y100 on loop A and W57 on the $\beta 2$ strand in loop D. The back walls are defined by W156 in loop B and L121 on the $\beta 6$ strand in loop E. The front wall of the pocket is formed by loop C, which packs tightly onto the ligand, contributing interactions from the vicinal cysteines and from Y197 and Y204. The hydrophobic top of the pocket is formed by V111 and F119 in loop E. In addition to the aromatic and hydrophobic interactions with these side chains, nicotine is poised to form a hydrogen bond between its electropositive pyrrolidine nitrogen and the backbone carbonyl oxygen of W156. The pyrrolidine nitrogen is also well oriented to form a cation– π interaction with the indole ring of W156, a recurring ligand–receptor interaction in the superfamily, although not always to this tryptophan²⁰. Residues in loop F do not contribute directly to nicotine binding; however, D170 on loop F probably stabilizes loop C via a hydrogen bond to the backbone nitrogen of C199 (Extended Data Figs 5 and 6).

To date, all high-resolution structural information for Cys-loop receptors has come from homopentameric assemblies, leaving many questions unanswered regarding architecture of the non-canonical interfaces. The $\alpha 4\beta 2$ crystal structure reveals a surprising reorganization of the conserved aromatic residues in the β – β and β – α interfaces that precludes nicotine binding. The source of the reorganization appears to be the identity of the residue that precedes the loop B tryptophan by two positions. In the $\alpha 4$ subunit, this residue is a glycine (G154); in $\beta 2$, it is an arginine (R149). When the $\beta 2$ subunit contributes to the (+) side of the interface (Fig. 2d, e), this R149 orients longitudinally into the base of the binding pocket. The second tyrosine on loop C is not present in the $\beta 2$ subunit, which allows Y196 to change its rotameric position, orienting towards the membrane. A second tyrosine, Y95 in loop A, rotates away from the membrane. The result of the switch in conformations of these two tyrosines is that the positively charged guanidinium group of R149 is sandwiched between their two aromatic rings, in a

sense satisfying the electron-rich π system as the pyrrolidine nitrogen of nicotine does in the α – β interfaces. A consequence of the reorganization around the arginine is that W151 in loop B must move; its side chain rotates out of the binding pocket completely. The conformations of these residues on the (+) side are similar between the β – β and β – α interfaces; the differences between them arise from the (-) side of the interface, where three hydrophobic groups on the (-) side of the $\beta 2$ subunit are replaced by polar side chains on the (-) side of the $\alpha 4$ subunit (Fig. 2e). This difference in chemical environment may affect nicotine binding to $\alpha 4$ – $\alpha 4$ interfaces in the 3 α :2 β stoichiometry²¹. The polar environment on the (-) face of the $\alpha 4$ subunit may be less favourable for nicotine binding in the orientation we observe at the α – β interfaces, wherein the pyridine ring packs against the hydrophobic (-) face of the β subunit. By comparison, the homopentameric $\alpha 7$ nicotinic receptor preserves two of the three hydrophobic residues in loop E (Extended Data Fig. 6a) and maintains nicotine binding, albeit with lower affinity.

After prolonged exposure to agonist, nicotinic receptors desensitize, adopting a high-affinity and agonist-bound, non-conducting conformation⁷. We performed patch-clamp electrophysiology experiments comparing responses of full-length and crystallized $\alpha 4\beta 2$ receptor constructs to acetylcholine and found them to behave similarly (Fig. 3a). We next measured responses to 1 mM nicotine, as was used throughout purification and for crystallization, and observed that the receptor desensitized profoundly within a few milliseconds. This functional result indicates that we would observe a desensitized, non-conducting conformation in the structure. The receptor structure reveals the transmembrane channel tapering to a constriction point at the interface with the cytosol (Fig. 3b). The narrowest point in the pore is defined by glutamate side chains at the -1' position of the M2 α -helices, which give rise to a constriction of 3.8 Å in diameter (Fig. 3b, c). The consensus on minimum pore diameter among cation-selective Cys-loop receptors is in the range of ~ 6 –8 Å (refs 22, 23), consistent with the permeant ion being at least partially hydrated. The $\alpha 4\beta 2$ receptor is a non-selective cation channel, being permeable

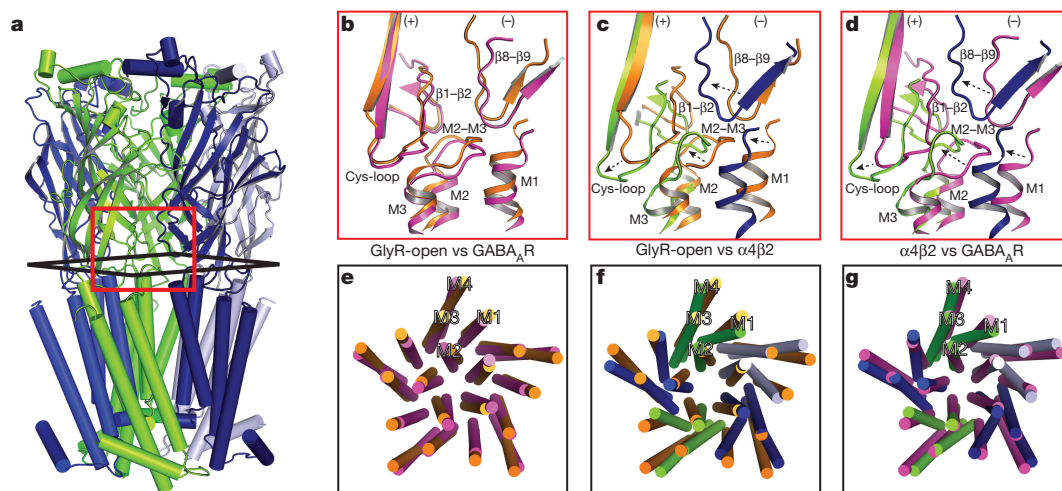


Figure 4 | Rearrangements at the membrane interface underlie desensitization in the $\alpha 4\beta 2$ receptor. **a**, Reference orientation of the $\alpha 4\beta 2$ receptor. **b–d**, Superimpositions of whole pentamers based on alignment of transmembrane domains, showing local structural differences at the membrane interface. **e–g**, Superimpositions of whole

pentamers based on alignment of extracellular domains, showing global differences in transmembrane domains. **b, e**, GlyR-open (orange) versus GABA_AR (magenta). **c, f**, GlyR-open versus $\alpha 4\beta 2$ structure (green, blue). **d, g**, $\alpha 4\beta 2$ versus GABA_AR.

to Na⁺, K⁺ and Ca²⁺. Na⁺ is the smallest, with an ionic diameter of 1.90 Å. Adding a single equatorial water molecule (2.8 Å diameter) would put the diameter of the permeant species above the observed constriction size. We compared the $\alpha 4\beta 2$ receptor pore conformation to those from recent structures that probably represent the three principal receptor states: resting/closed (glycine receptor plus strychnine²⁴; GlyR-closed), activated/open (glycine receptor plus glycine²⁴; GlyR-open) and desensitized/closed (GABA_AR¹⁷) (Fig. 3c and Extended Data Fig. 7). The pore conformation of the $\alpha 4\beta 2$ receptor most closely resembles the desensitized GABA_AR, where the gate is at the cytosolic end of the pore. Functional studies also suggest that the desensitization gate is located at the cytosolic side of the pore²⁵. Thus, structural and functional analyses are consistent with the $\alpha 4\beta 2$ receptor structure representing a desensitized, non-conducting state.

To probe mechanisms of ion selectivity, we analysed the electrostatic properties of the permeation pathway of the $\alpha 4\beta 2$ receptor (Fig. 3d). The surface of the extracellular vestibule is strongly electronegative, which probably serves to increase the local concentration of cations near the channel mouth. The electrostatic potential becomes more neutral at the extracellular end of the pore, where the 20' glutamate side chains from the two $\alpha 4$ subunits are offset by the 20' lysine side chains from the three $\beta 2$ subunits. This 20' position is the only site in the pore where the $\alpha 4$ and $\beta 2$ subunits contribute opposing charges to the electrostatic surface, and thus is where alternate subunit stoichiometries would be expected to influence permeation properties most strongly. Indeed, the higher Ca²⁺ permeability of the 3 α :2 β stoichiometry of this receptor has been shown to result from the swap of lysine to glutamate at the 20' position in that assembly⁹. Approaching the constriction point in the pore, the surface becomes strongly electronegative, dominated by the five glutamate side chains that form the selectivity filter at the base of the pore. The side chains are folded towards the pore axis with their carboxylates probably stabilized through hydrogen bonding with the –2' backbone carbonyl oxygens from adjacent subunits.

To move beyond the local conformation observed in the pore, and to place the $\alpha 4\beta 2$ receptor structure in the context of the resting–activated–desensitized gating cycle, we next compared the overall conformation of the $\alpha 4\beta 2$ receptor to the reference structures for distinct conformations. Structures of GluCl²⁶ and the glycine receptor²⁴, each in multiple conformations, suggest that within an individual subunit, the extracellular (ECDs) and transmembrane subdomains (TMDs) behave in large part as rigid bodies during state transitions. Thus we initially compared the ECD and TMD of an $\alpha 4$ subunit with the analogous

subdomains from the open and desensitized structures described earlier (Extended Data Fig. 8a–c). We found that the C α backbones from these subdomains superimpose well (C α root mean squared deviation (r.m.s.d.) 1.6–2.8 Å), with noteworthy differences in loops at the extracellular–transmembrane interface thought to be involved in signal transduction. These loops include the $\beta 1$ – $\beta 2$, M2–M3 and Cys-loops from the (+) subunit and the $\beta 8$ – $\beta 9$ loop and the $\beta 10$ –M1 helix junction in the (–) subunit. To understand how the reorganization of these interfacial loops relates to global conformational changes, we superimposed whole receptors based on alignment of their pentameric transmembrane domains, and examined corresponding differences in the extracellular domains. We were surprised to find that while the GABA_AR pore is tightly closed, more so even than $\alpha 4\beta 2$ (Fig. 3c), the conformation of the GABA_AR extracellular domain much more closely resembles the open GlyR structure than the $\alpha 4\beta 2$ receptor structure (Extended Data Fig. 8d, e).

Examination of the interactions between the extracellular and transmembrane domains further illustrates the differences between the open and the two desensitized conformations (Fig. 4a–d). At the ECD–TMD interface, local loop conformations are similar between the GlyR-open and the GABA_AR structures (Fig. 4b). Comparison of $\alpha 4\beta 2$ with both the GlyR-open (Fig. 4c) and the GABA_AR (Fig. 4d) structures reveals concerted displacements in $\alpha 4\beta 2$ of the $\beta 1$ – $\beta 2$, M2–M3 and Cys-loops on the (+) subunit and the $\beta 8$ – $\beta 9$ loop and the $\beta 10$ –M1 helix on the (–) subunit. These displacements are maximal at the Cys-loop, with differences between reference C α atoms of 6.5 Å for $\alpha 4\beta 2$ versus GABA_AR and 7.4 Å for $\alpha 4\beta 2$ versus GlyR-open. Analysis of the conformational differences at the subunit level between $\alpha 4\beta 2$ and GlyR-open that generate these displacements suggests a 15° rotation around an axis passing through the Cys-loop (Extended Data Fig. 8f). This rotation results in closure of the ion channel and necessitates reorganization of the ECD–TMD interface. In contrast, analysis of the conformational differences between $\alpha 4\beta 2$ and GABA_AR suggests a 13° tilting of the ECD (Extended Data Fig. 8g). As a result, from $\alpha 4\beta 2$ to the GABA_AR, the pore remains similarly closed, but the ECD–TMD interface is different. In both cases, the resulting displacement of the Cys-loop at the pivot point coincides with a major alteration in the conformation of the M1 helix of $\alpha 4\beta 2$ relative to GlyR-open and to GABA_AR (Fig. 4e–g).

Our structural analysis suggests that the $\alpha 4\beta 2$ and GABA_AR structures represent distinct desensitized states. Kinetically distinct desensitized states are well described for both GABA_A and nicotinic

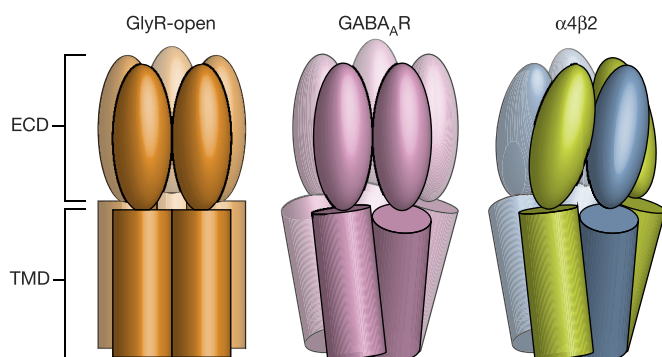


Figure 5 | Conformational changes underlying desensitization. Cartoon illustrates the relative positions of ECDs and TMDs in the $\alpha 4\beta 2$ receptor compared to the open conformation of the glycine receptor and the desensitized conformation of the GABA_A receptor.

receptors^{27,28}. The electrophysiology data for nicotine at the $\alpha 4\beta 2$ receptor, and other studies of nicotine at the rat $\alpha 4\beta 2$ receptor²⁹, are consistent with a desensitized receptor; those presented with the GABA_AR structure are potentially consistent with an intermediate or transitional state stabilized by the novel agonist benzamidine. We speculate that the extensive conformational rearrangements observed in the $\alpha 4\beta 2$ receptor ECD–TMD interface further stabilize the receptor and thereby contribute to the increased affinity for agonist in the desensitized state⁷. This progression of quaternary rearrangements is illustrated in Fig. 5. These interpretations are tentative as both of these structures were determined in the presence of detergent, removed from the native membrane environment known to be important for pentameric receptor function³⁰. Additional Cys-loop receptor structures in desensitized states, and nicotinic receptor structures in additional states, will help elucidate the detailed structural changes underlying desensitization.

Here we describe the X-ray structure of a nicotinic acetylcholine receptor, the heteropentameric $\alpha 4\beta 2$ receptor. This structure of a heteromeric Cys-loop receptor sheds light on the architecture of the neurotransmitter site with bound nicotine and illustrates why the two other classes of binding sites are unable to bind classical nicotinic agonists. The receptor is locked in a non-conducting, desensitized conformation by the agonist nicotine. The $\alpha 4\beta 2$ receptor conformation is distinct from prior structural information on a desensitized GABA_A receptor, and thereby provides an important addition towards mapping the structural basis of allosteric gating in Cys-loop receptors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 April; accepted 22 August 2016.

Published online 3 October 2016.

- Langley, J. N. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *J. Physiol. (Lond.)* **33**, 374–413 (1905).
- Changeux, J. P. & Edelstein, S. J. *Nicotinic Acetylcholine Receptors: From Molecular Biology to Cognition* (Odile Jacob Publishing Corporation, 2005).
- Engel, A. G., Shen, X. M., Selcen, D. & Sine, S. M. Congenital myasthenic syndromes: pathogenesis, diagnosis, and treatment. *Lancet Neurol.* **14**, 420–434 (2015).
- Becchetti, A., Aracri, P., Meneghini, S., Brusco, S. & Amadeo, A. The role of nicotinic acetylcholine receptors in autosomal dominant nocturnal frontal lobe epilepsy. *Front. Physiol.* **6**, 22 (2015).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Rucktooa, P., Smit, A. B. & Sixma, T. K. Insight in nAChR subtype selectivity from AChBP crystal structures. *Biochem. Pharmacol.* **78**, 777–787 (2009).
- Nemecz, Á., Prevost, M. S., Menny, A. & Corring, P. J. Emerging molecular mechanisms of signal transduction in pentameric ligand-gated ion channels. *Neuron* **90**, 452–470 (2016).

- Nelson, M. E., Kuryatov, A., Choi, C. H., Zhou, Y. & Lindstrom, J. Alternate stoichiometries of $\alpha 4\beta 2$ nicotinic acetylcholine receptors. *Mol. Pharmacol.* **63**, 332–341 (2003).
- Tapia, L., Kuryatov, A. & Lindstrom, J. Ca²⁺ permeability of the $(\alpha 4)_3(\beta 2)_2$ stoichiometry greatly exceeds that of $(\alpha 4)_2(\beta 2)_3$ human acetylcholine receptors. *Mol. Pharmacol.* **71**, 769–776 (2007).
- Lester, H. A. et al. Nicotine is a selective pharmacological chaperone of acetylcholine receptor number and stoichiometry. Implications for drug discovery. *AAPS J.* **11**, 167–177 (2009).
- Morales-Perez, C. L., Novello, C. M. & Hibbs, R. E. Manipulation of subunit stoichiometry in heteromeric membrane proteins. *Structure* **24**, 797–805 (2016).
- Zwart, R. et al. 5-I A-85380 and TC-2559 differentially activate heterologously expressed $\alpha 4\beta 2$ nicotinic receptors. *Eur. J. Pharmacol.* **539**, 10–17 (2006).
- Zhou, Y. et al. Human $\alpha 4\beta 2$ acetylcholine receptors formed from linked subunits. *J. Neurosci.* **23**, 9004–9015 (2003).
- Hassaine, G. et al. X-ray structure of the mouse serotonin 5-HT₃ receptor. *Nature* **512**, 276–281 (2014).
- Karlin, A. Emerging structure of the nicotinic acetylcholine receptors. *Nature Rev. Neurosci.* **3**, 102–114 (2002).
- Sauguet, L. et al. Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO J.* **32**, 728–741 (2013).
- Miller, P. S. & Aricescu, A. R. Crystal structure of a human GABA_A receptor. *Nature* **512**, 270–275 (2014).
- Picciotto, M. R. et al. Acetylcholine receptors containing the $\beta 2$ subunit are involved in the reinforcing properties of nicotine. *Nature* **391**, 173–177 (1998).
- Tapper, A. R. et al. Nicotine activation of $\alpha 4^*$ receptors: sufficient for reward, tolerance, and sensitization. *Science* **306**, 1029–1032 (2004).
- Dougherty, D. A. The cation– π interaction. *Acc. Chem. Res.* **46**, 885–893 (2013).
- Mazzaferro, S. et al. Additional acetylcholine (ACh) binding site at $\alpha 4/\alpha 4$ interface of $(\alpha 4\beta 2)_2\alpha 4$ nicotinic receptor influences agonist sensitivity. *J. Biol. Chem.* **286**, 31043–31054 (2011).
- Yang, J. Ion permeation through 5-hydroxytryptamine-gated channels in neuroblastoma N18 cells. *J. Gen. Physiol.* **96**, 1177–1198 (1990).
- Dwyer, T. M., Adams, D. J. & Hille, B. The permeability of the endplate channel to organic cations in frog muscle. *J. Gen. Physiol.* **75**, 469–492 (1980).
- Du, J., Lü, W., Wu, S., Cheng, Y. & Gouaux, E. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **526**, 224–229 (2015).
- Gielen, M., Thomas, P. & Smart, T. G. The desensitization gate of inhibitory Cys-loop receptors. *Nature Commun.* **6**, 6829 (2015).
- Althoff, T., Hibbs, R. E., Banerjee, S. & Gouaux, E. X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors. *Nature* **512**, 333–337 (2014).
- Celentano, J. J. & Wong, R. K. Multiphasic desensitization of the GABA_A receptor in outside-out patches. *Biophys. J.* **66**, 1039–1050 (1994).
- Léna, C. & Changeux, J. P. Allosteric modulations of the nicotinic acetylcholine receptor. *Trends Neurosci.* **16**, 181–186 (1993).
- Paradiso, K. G. & Steinbach, J. H. Nicotine is highly effective at producing desensitization of rat $\alpha 4\beta 2$ neuronal nicotinic receptors. *J. Physiol.* **553**, 857–871 (2003).
- Labriola, J. M. et al. Structural sensitivity of a prokaryotic pentameric ligand-gated ion channel to its membrane environment. *J. Biol. Chem.* **288**, 11294–11303 (2013).

Acknowledgements We thank J. Lindstrom for providing the $\alpha 4$ and $\beta 2$ receptor genes, D. Borek and Z. Otwinowski for guidance in structural analyses, J. Cabrera for assistance with illustrations, and members of the Hibbs laboratory for comments on the manuscript. X-ray diffraction experiments at the Argonne National Laboratory's Advanced Photon Source 24-ID-C beamline were supported by the National Institutes of Health (NIH; GM103403 and RR029205) and the Department of Energy (DE-AC02-06CH11357). This research project was supported by an NIH training grant (T32 NS069562) and a Howard Hughes Medical Institute Gilliam Fellowship to C.L.M.-P.; R.E.H. is supported by a McKnight Scholar Award, a Klingenstein-Simons Fellowship Award in the Neurosciences, The Welch Foundation (I-1812), The Friends of the Alzheimer's Disease Center and the NIH (DA037492, DA042072 and NS077983).

Author Contributions C.L.M.-P., C.M.N. and R.E.H. contributed to all aspects of the project.

Author Information Atomic coordinates and structure factors for the $\alpha 4\beta 2$ plus nicotine structure have been deposited in the Protein Data Bank under accession 5KX1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.E.H. (ryan.hibbs@utsouthwestern.edu).

Reviewer Information Nature thanks M. Akabas, R. Dutzler and A. Karlin for their contribution to the peer review of this work.

METHODS

Protein expression and purification. The human $\alpha 4$ and $\beta 2$ nicotinic receptor genes were provided by J. Lindstrom at the University of Pennsylvania. For the purposes of small-scale biochemical screening, a synthesized gene encoding enhanced green fluorescent protein (eGFP) was spliced into the M3–M4 loop of each subunit and the genes were subcloned into the pEZT bacmam expression vector¹¹. The eGFP fusion to one subunit was co-transfected into GnTI- HEK cells (ATCC CRL-3022) with a panel of deletion constructs for the partner subunit; a large number of constructs were screened in this manner for expression and pentameric monodispersity by fluorescence-detection size-exclusion chromatography (FSEC)³¹. The final expression constructs for crystallization included the native signal peptides and residues 1–338 and 556–601 in the $\alpha 4$ subunit and residues 1–330 and 417–477 in the $\beta 2$ subunit (residue numbering here is for the wild-type mature, signal-peptide-cleaved protein sequence). Deletion of the M3–M4 loop has been shown to not affect function in other Cys-loop receptor family members³². To promote crystallization a Glu-Arg linker was inserted in the MX–M4 junction, between Phe559–Ser560 in the $\alpha 4$ subunit and between Gln420–Ser421 in the $\beta 2$ subunit. For purification purposes a Strep-tag was inserted at the C terminus of the $\beta 2$ subunit preceded by a Ser–Ala linker. Previously identified expression conditions resulted in a homogenous receptor subunit stoichiometry of two $\alpha 4$ and three $\beta 2$ subunits¹¹. For large-scale expression, 1.6 l of suspension GnTI- cells were transduced with multiplicities of infection (MOIs) of 0.25:0.5 for the $\alpha 4$ and $\beta 2$ subunits, respectively. Nicotine (Sigma-Aldrich) and sodium butyrate (Sigma-Aldrich) were added at the time of transduction to 0.1 mM and 3 mM, respectively. At the time of transduction, suspension cells were moved to 30 °C and 8% CO₂. After 72 h, cells were collected by centrifugation, resuspended in 20 mM Tris, pH 7.4, 150 mM NaCl (TBS buffer), 1 mM nicotine and 1 mM phenylmethanesulfonyl fluoride (Sigma-Aldrich), and disrupted using an Avestin Emulsiflex. Lysed cells were centrifuged for 15 min at 10,000g; supernatants containing membranes were centrifuged 2 h at 186,000g. Membrane pellets were mechanically homogenized and solubilized for 1 h at 4 °C, in a solution containing TBS, 40 mM *n*-dodecyl- β -D-maltopyranoside (DDM; Anatrace), 1 mM nicotine and 0.2 mM cholesteryl hemisuccinate (CHS; Anatrace). Solubilized membranes were centrifuged for 40 min at 186,000 g then passed over high-capacity Strep-Tactin (IBA) affinity resin. The resin was washed with size-exclusion chromatography (SEC) buffer containing TBS, 1 mM DDM, 1 mM nicotine, 0.2 mM CHS and 1 mM TCEP (Thermo Fisher Scientific) and eluted in the same buffer containing 5 mM desthiobiotin (Sigma-Aldrich). Peak elution fractions were concentrated and digested with Endoglycosidase H overnight in a 1:8 w:w ratio at 4 °C. This material was then injected over a Superose 6 10/300 GL column equilibrated in SEC buffer wherein DDM was replaced with 2 mM *n*-undecyl- β -D-maltopyranoside (Anatrace). Peak fractions were assayed by FSEC, monitoring tryptophan fluorescence, before pooling and concentrating for crystallization. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Crystallization, X-ray data collection and structure solution. Purified $\alpha 4\beta 2$ was concentrated to 1.5–2.5 mg/ml in SEC buffer and crystallized by hanging-drop vapour diffusion. The best diffracting crystals of the nicotine-bound receptor were obtained after mixing protein with reservoir solution containing 0.05 M ADA pH 6.8, 12.5% PEG 1500 and 10% PEG 1000 in a 1:1 ratio and incubating over sealed wells containing 0.5 ml reservoir, at 14 °C. The crystals were cryoprotected with additional PEG 1000, PEG 1500 and ethylene glycol before flash freezing in liquid nitrogen. Crystals of the 5-Iodo-A-85380 (IA)³³-bound receptor were obtained using the same approach, however, the protein was purified in the absence of ligand, with IA added after SEC to a concentration of 0.5 mM. The best-diffracting crystals of the IA complex were obtained at 14 °C using a reservoir solution of 0.05 M ADA pH 6.5 and 24% PEG 400; crystals were cryoprotected with additional PEG 400 before flash freezing in liquid nitrogen. X-ray data were collected at the 24-ID-C beamline at the Advanced Photon Source (Argonne, IL). Both data sets were collected from single crystals. The data set from the IA complex was collected at low energy (7,300 eV) to maximize anomalous signal from iodine in the ligand.

Diffraction data sets were integrated and scaled using HKL2000³⁴. The 'auto corrections' option was used to assess anisotropic signal to noise, determine the resolution to use in refinement, and perform ellipsoidal truncation of the data as well as anisotropic B factor sharpening. The data from the nicotine complex were highly anisotropic, extending to ~ 3.6 Å in the best direction and ~ 4.5 Å in the worst. Electron density maps using the auto-corrected data contain far more features than the unmodified data and thus were used for all of the manual model building. However, truncated data from 'auto corrections' suffer from low completeness in the high-resolution shells. We thus used the UCLA diffraction anisotropy server³⁵ to perform more conservative truncation and sharpening of the data; the deposited model underwent a final round of refinement against this

truncated data set to generate the statistics shown in Extended Data Table 1. The deposited structure factors include both sets of these truncated, sharpened data.

The structure of the nicotine-bound $\alpha 4\beta 2$ receptor was solved by molecular replacement using a pentameric homology model based on the desensitized GABA_A $\beta 3$ receptor structure (PDB accession 4COF)¹⁷, with models of the acetylcholine receptor $\alpha 4$ and $\beta 2$ subunits generated using Swissmodel³⁶. A panel of homology models was made comprising different orderings of subunits around the pentameric ring; the best molecular replacement search model had an ordering of α - β - β - α - β . Distinct electron density features, mainly in loop C, provided the first convincing clues into subunit identity. Swapping positions of $\alpha 4$ and $\beta 2$ subunits in the pentamer, followed by monitoring of *R* factors after refinement, supported the subunit assignment; however, we sought additional validation. The potent agonist IA is expected to bind only in the canonical neurotransmitter site found at α - β interfaces. We exploited anomalous signal in a low-resolution data set of the $\alpha 4\beta 2$ -IA complex to independently validate subunit assignment. After rigid body refinement of the nicotine-bound model in this IA-complex data set, strong anomalous difference peaks were observed: one in each of the two binding pockets that we had assigned as $\alpha 4$ - $\beta 2$ interfaces (4.5σ and 5.8σ) and similarly strong peaks near Cys-loop disulfides where four sulfur atoms are in close proximity. No anomalous difference signal was observed at the corresponding position in the β - α or β - β interfaces. Once the subunit arrangement was confirmed, iterative cycles of manual rebuilding in Coot³⁷, jelly body refinement in Refmac³⁸ and further restrained refinement in Phenix³⁹ were performed. The Fitmunk server⁴⁰ was used to identify improved side chain rotamers. Torsion-angle non-crystallographic symmetry restraints ($\alpha 4$ subunits and $\beta 2$ subunits as separate groups), group *B* factors (one per residue) and TLS parameters (two groups per subunit) were used in refinement with Phenix.

The ECDs and TMDs were modelled with a high degree of confidence, with electron density visible for most side chains, one GlcNAc residue per subunit and two molecules of nicotine. One exception to the overall well-ordered ECD is the distal end of loop C in the $\beta 2$ subunits, which exhibited weak electron density in two of the three β subunits, and thus its modelling is tentative. A pancake-shaped difference electron density peak midway along the ion channel was modelled as a sodium ion coordinated by water molecules mediating hydrogen bonds to the proximal threonine side chains. The sodium ion and water assignments are speculative; they were based on NaCl being the only salt present in purification and crystallization, the channel being selective for cations, *B* factors after refinement, and a similar arrangement of sodium and water in the high-resolution structure of the bacterial pH-gated cation channel GLIC¹⁶. The register matches that of the AChBPs in the extracellular domain and the 5-HT₃R, GABA_AR, GlyR and GluCl structures in the transmembrane domain. Comparisons were also made with the *Torpedo* ACh receptor structure and were found to be different in register throughout much of the TMD, as previously described^{17,41–43}. There was no observable electron density for 7 residues in the N terminus of the $\alpha 4$ subunit, 11 and 15 residues linking the MX helix (following M3) to the M4 helix of the $\alpha 4$ and $\beta 2$ subunits and 5 and 30 residues from the C termini of the $\alpha 4$ and $\beta 2$ subunits. While there was clear electron density for the MX helix, the observable density between M3 and M4 was disordered relative to the rest of the receptor leading to some ambiguity in modelling, in particular in the linker between the M3 helix and the MX helix. In the final refined model the MX helix register matches that observed in the 5-HT₃R structure¹⁴. The five glutamate residues that define the pore constriction were not all well resolved. We modelled all five side chains in the same rotameric conformation based on convincing electron density for a subset. In the open state these glutamates are probably highly dynamic, with heterogeneous conformations affecting conductance⁴⁴.

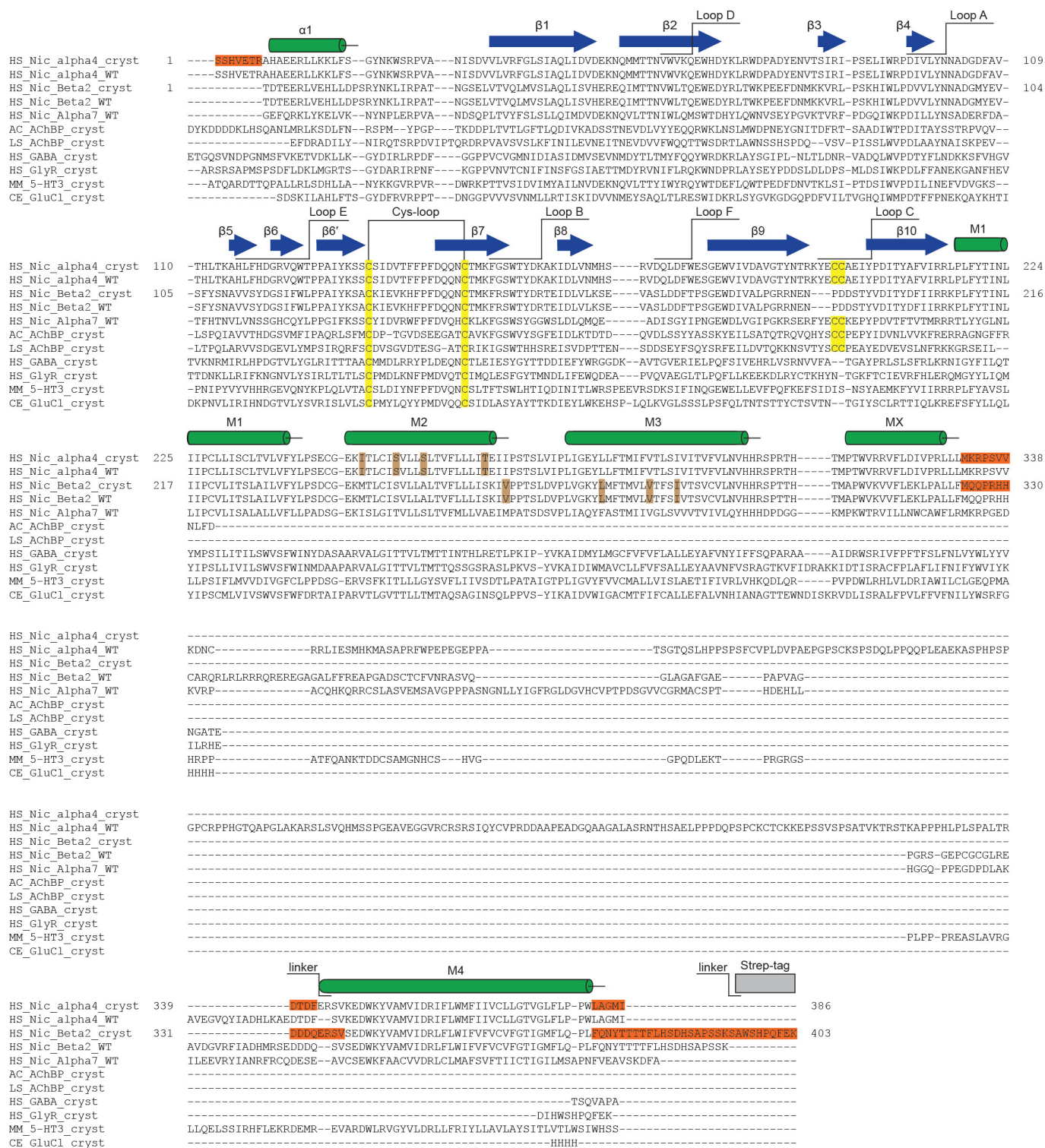
Sequence alignments were made using PROMALS3D⁴⁵. Ligand–receptor interactions were analysed with areaimol in the CCP4 suite^{46,47} and the CAPTURE program⁴⁸. Structural superpositions were made using Superpose⁴⁹ in the CCP4 suite. Subunit interfaces were analysed using the PDBe-PISA server⁵⁰. Pore diameters were calculated using HOLE⁵¹. Structural figures were made with PyMOL (Schrodinger, LLC) including the APBS electrostatics plugin⁵². Crystallographic software packages were compiled by SBGrid⁵³. Domain movements were analysed using DynDom (<http://fizz.cmp.uea.ac.uk/dyndom/>).

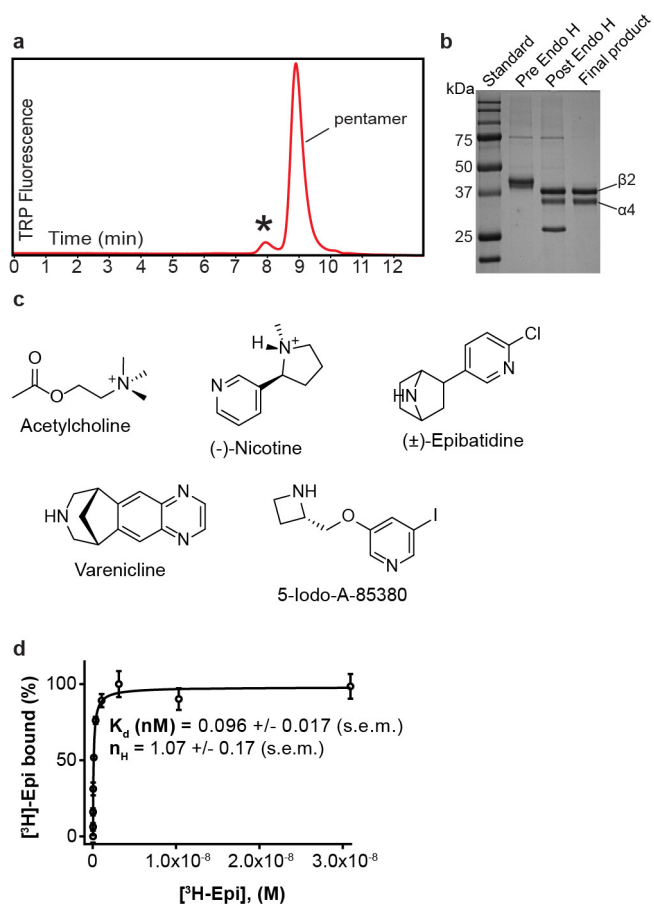
Radioligand binding. Experiments to measure binding of [³H]-epibatidine (PerkinElmer, 32.46 Ci/mmol) to the $\alpha 4\beta 2$ receptor, as well as competition with other ligands, were performed with protein purified as for crystallization but in the absence of ligands. The concentration of binding sites was kept at 0.1 nM after a preliminary experiment to determine optimal receptor concentration. In addition to the protein, the binding assay conditions included 20 mM Tris pH 7.4, 150 mM NaCl, 1 mM DDM, and 1 mg/ml streptavidin–YiSi scintillation proximity assay beads (SPA; GE Healthcare Life Sciences). Non-specific signal was determined in the presence of 100 μ M [¹H]-nicotine; all data shown are from

background-subtracted measurements. For competition assays [^3H]-epibatidine concentration was fixed at 1 nM. All data were analysed using Prism 6 software (GraphPad) with variable Hill slope. K_i values were calculated based on the experimentally determined K_d of 96 pM for [^3H]-epibatidine.

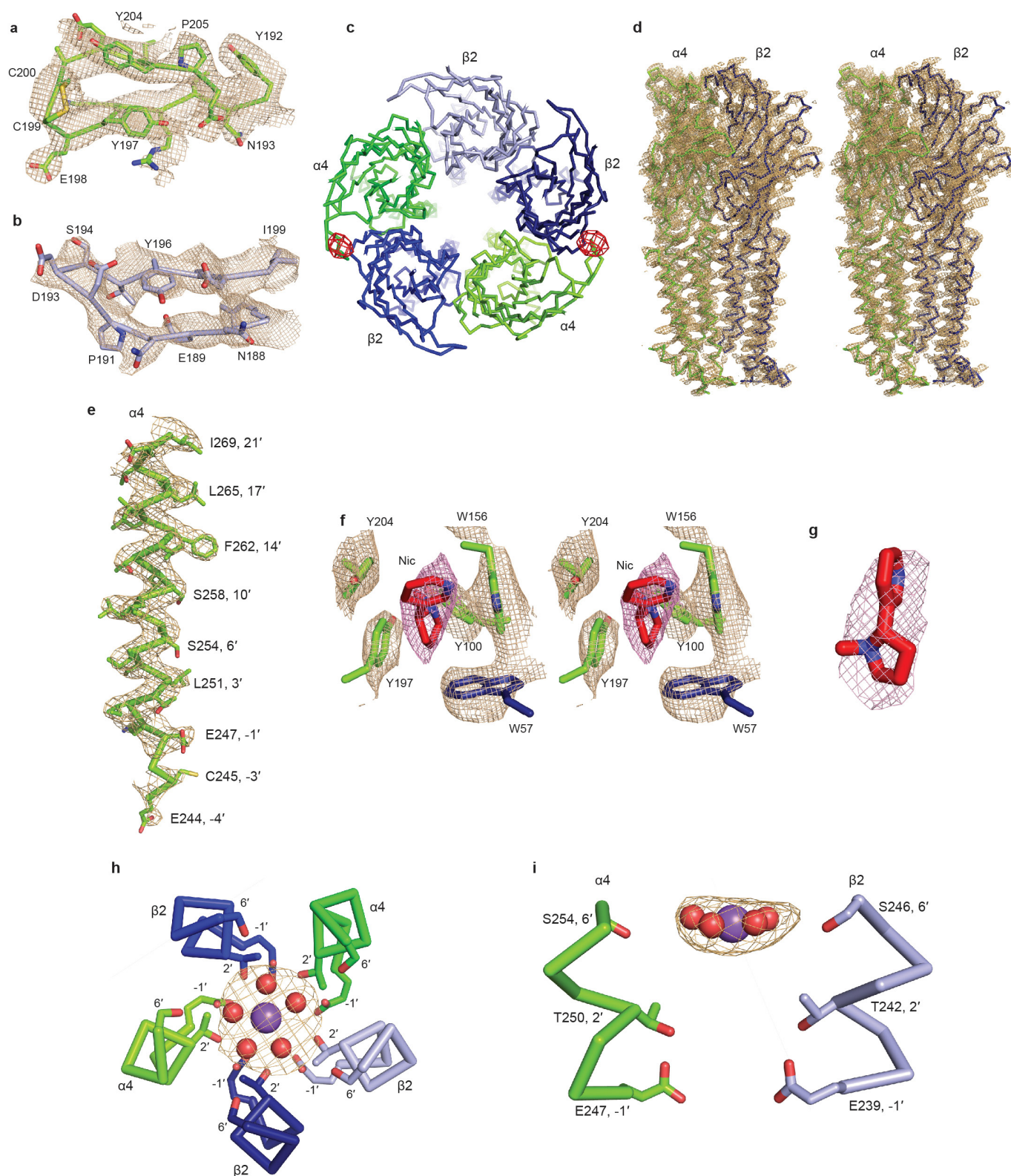
Electrophysiology. To test the $\alpha 4\beta 2$ receptor channel function, adherent GnTI-HEK cells were transfected with 0.5 μg of plasmid DNA for each subunit and 0.2 μg of a GFP expression plasmid using Lipofectamine 2000 (Thermo Fisher Scientific). The GFP expression plasmid was included to identify the cells for recording. After incubating for 72 h at 30 °C and 5% CO_2 the cells were patched using the whole-cell configuration and clamped at a membrane potential of -90 mV. The recordings were made with an Axopatch 200B amplifier, low-pass filtered at 5 kHz and digitized at 10 kHz using the Digidata 1440A and pClamp 10 software (Molecular Devices). Borosilicate glass pipettes (King Precision Glass) were pulled and polished to 2–4 M Ω resistance. The bath solution contained (in mM): 140 NaCl, 2.4 KCl, 4 CaCl_2 , 4 mgCl_2 , 10 HEPES pH 7.3 and 10 glucose. The pipette solution contained (in mM): 150 CsF, 10 NaCl, 10 EGTA, 20 HEPES pH 7.3. The acetylcholine chloride (Sigma-Aldrich) and nicotine solutions were prepared in bath solution. Solution exchange was achieved using a gravity driven RSC-200 rapid solution changer (Bio-Logic).

31. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
32. Jansen, M., Bali, M. & Akabas, M. H. Modular design of Cys-loop ligand-gated ion channels: functional 5-HT $_3$ and GABA $\rho 1$ receptors lacking the large cytoplasmic M3M4 loop. *J. Gen. Physiol.* **131**, 137–146 (2008).
33. Mukhin, A. G. *et al.* 5-Iodo-A-85380, an $\alpha 4\beta 2$ subtype-selective ligand for nicotinic acetylcholine receptors. *Mol. Pharmacol.* **57**, 642–649 (2000).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
36. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
38. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
40. Porebski, P. J., Cymborowski, M., Pasenkiewicz-Gierula, M. & Minor, W. Fitmunk: improving protein structures by accurate, automatic modeling of side-chain conformations. *Acta Crystallogr. D* **72**, 266–280 (2016).
41. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* **474**, 54–60 (2011).
42. Mnatsakanyan, N. & Jansen, M. Experimental determination of the vertical alignment between the second and third transmembrane segments of muscle nicotinic acetylcholine receptors. *J. Neurochem.* **125**, 843–854 (2013).
43. Corringer, P. J. *et al.* Atomic structure and dynamics of pentameric ligand-gated ion channels: new insight from bacterial homologues. *J. Physiol.* **588**, 565–572 (2010).
44. Cymes, G. D. & Grosman, C. The unanticipated complexity of the selectivity-filter glutamates of nicotinic receptors. *Nature Chem. Biol.* **8**, 975–981 (2012).
45. Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.* **1079**, 263–271 (2014).
46. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
47. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
48. Gallivan, J. P. & Dougherty, D. A. Cation- π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
49. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).
50. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
51. Smart, O. S., Neduelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).
52. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
53. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
54. Hibbs, R. E. *et al.* Structural determinants for interaction of partial agonists with acetylcholine binding protein and neuronal $\alpha 7$ nicotinic acetylcholine receptor. *EMBO J.* **28**, 3040–3051 (2009).
55. Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
56. Huang, X., Chen, H., Michelsen, K., Schneider, S. & Shaffer, P. L. Crystal structure of human glycine receptor- $\alpha 3$ bound to antagonist strychnine. *Nature* **526**, 277–280 (2015).
57. Wonnacott, S. & Barik, J. Nicotinic ACh receptors. *Tocris Reviews* **28**, 1–20 (2007).
58. Gonzalez-Gutierrez, G. & Grosman, C. Bridging the gap between structural models of nicotinic receptor superfamily ion channels and their corresponding functional states. *J. Mol. Biol.* **403**, 693–705 (2010).
59. Parikh, R. B., Bali, M. & Akabas, M. H. Structure of the M2 transmembrane segment of GLIC, a prokaryotic Cys loop receptor homologue from *Gloeobacter violaceus*, probed by substituted cysteine accessibility. *J. Biol. Chem.* **286**, 14098–14109 (2011).
60. Laha, K. T., Ghosh, B. & Czajkowski, C. Macroscopic kinetics of pentameric ligand gated ion channels: comparisons between two prokaryotic channels and one eukaryotic channel. *PLoS ONE* **8**, e80322 (2013).
61. Bouzat, C., Bartos, M., Corradi, J. & Sine, S. M. The interface between extracellular and transmembrane domains of homomeric Cys-loop receptors governs open-channel lifetime and rate of desensitization. *J. Neurosci.* **28**, 7808–7819 (2008).
62. Fourati, Z., Sauguet, L. & Delarue, M. Genuine open form of the pentameric ligand-gated ion channel GLIC. *Acta Crystallogr. D* **71**, 454–460 (2015).



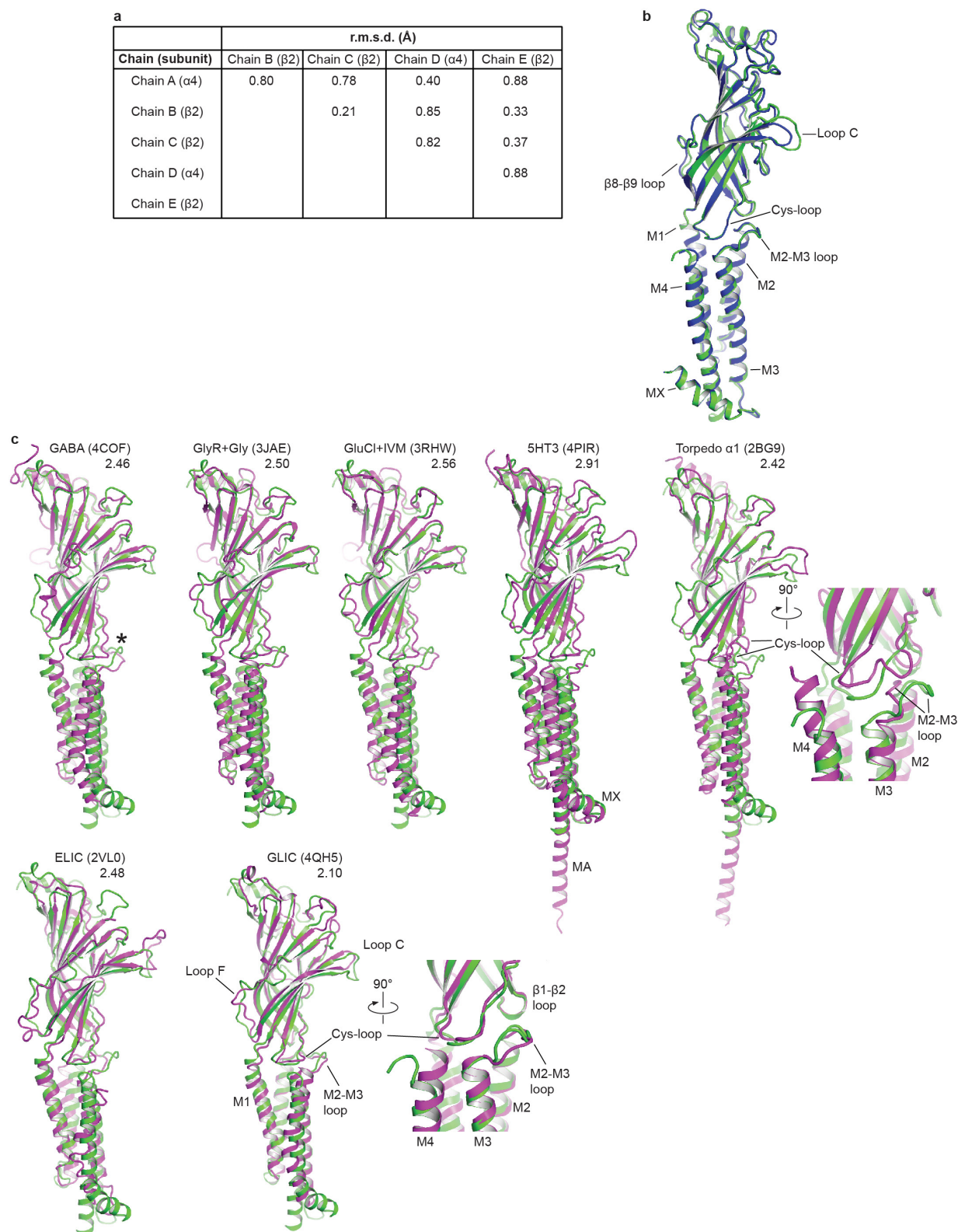


Extended Data Figure 2 | Biochemical analysis. **a**, Fluorescence-detection size-exclusion chromatography (FSEC) trace of the $\alpha 4\beta 2$ nicotinic receptor. The protein sample used for crystallization was tested by FSEC using an SRT SEC-500 column (0.35 ml min^{-1}) monitoring tryptophan fluorescence. The receptor exhibited time-dependent oligomerization/aggregation indicated by an asterisk. Pentamer indicates the elution peak of the heteropentameric assembly. **b**, SDS-polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie of the stages of receptor purification. **c**, Chemical structures of ligands used in crystallization, electrophysiology and binding assays. **d**, Saturation binding experiments with $[^3\text{H}]$ -epibatidine. Binding affinity (K_d) was calculated using the one site binding with variable slope equation in Graphpad Prism. The published range for epibatidine K_b , for reference, is $0.042\text{--}0.150 \text{ nM}$ (all published values are from a pharmacological review⁵⁷). The experiment was performed in triplicate. n_H , Hill coefficient. Error bars are s.e.m.



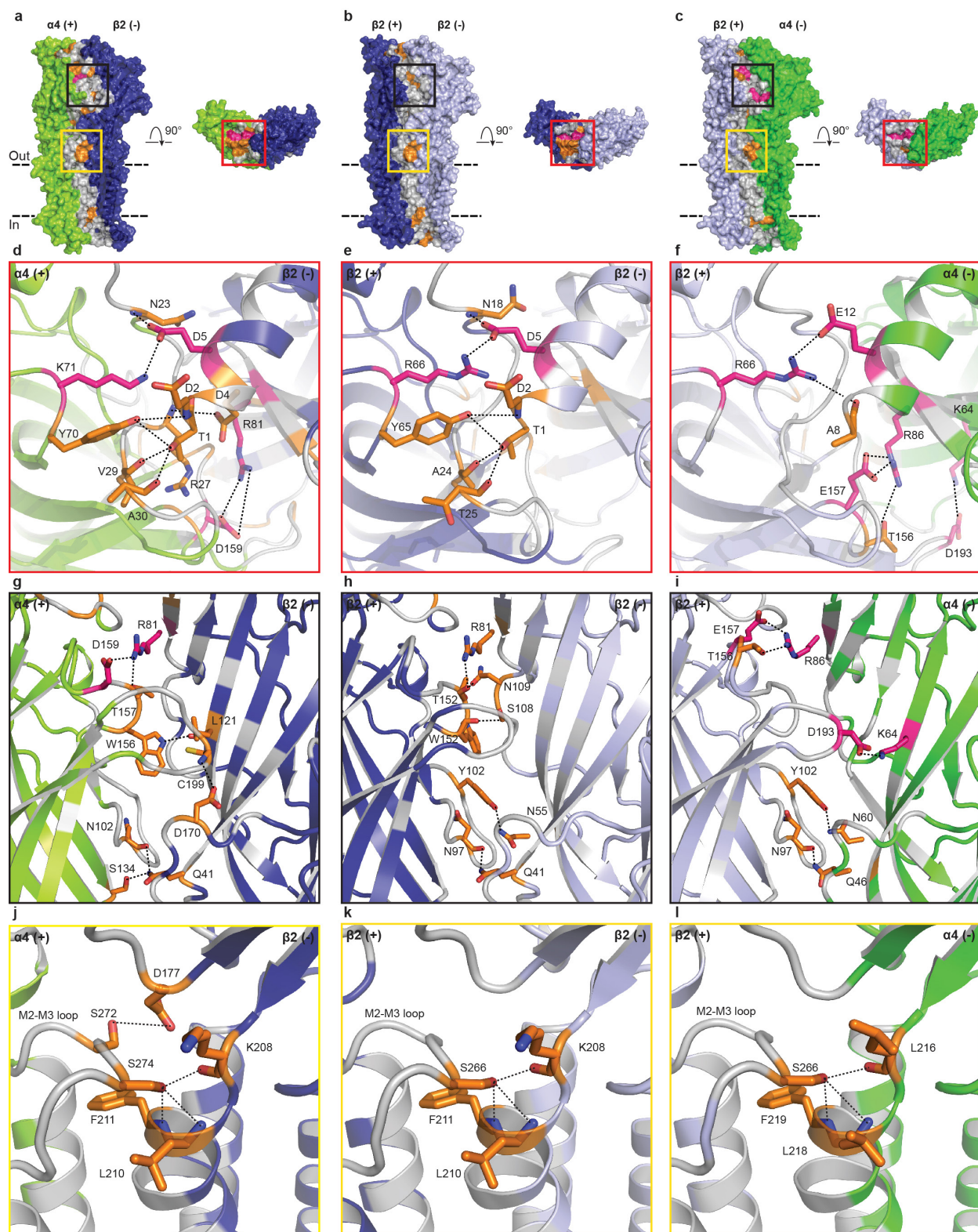
Extended Data Figure 3 | Electron density quality. **a, b**, $2F_o - F_c$ electron density maps of loop C from an $\alpha 4$ and $\beta 2$ subunit, respectively (contoured at 1σ), with reference residues indicated. Perspective is from inside binding pocket looking towards receptor periphery. **c**, View down the channel axis towards the cytosol. Anomalous difference peaks from co-crystallization with 5-Iodo-A-85380 are shown as red mesh and contoured at 3σ . No detectable anomalous signal was present in other interfacial pockets. **d**, Stereo pair of $2F_o - F_c$ electron density maps (contoured at 1.5σ) from an interface of $\alpha 4$ and $\beta 2$ subunits. **e**, $F_o - F_c$

electron density map of an $\alpha 4$ subunit M2 α -helix (contoured at 1.5σ). Reference residues in the M2 helix are indicated. **f**, Stereo pair of $F_o - F_c$ omit maps (contoured at 2σ) of selected residues and nicotine in the neurotransmitter-binding pocket. Residues and ligand omitted from map calculation are labelled. **g**, $F_o - F_c$ omit map (contoured at 2σ) for nicotine in the α - β interface. **h-i**, $F_o - F_c$ omit map (contoured at 2σ) of the ion and waters in the pore. The Na^+ ion (purple) and water (red) are represented as spheres. The nearest residues on the M2 α -helices are indicated.



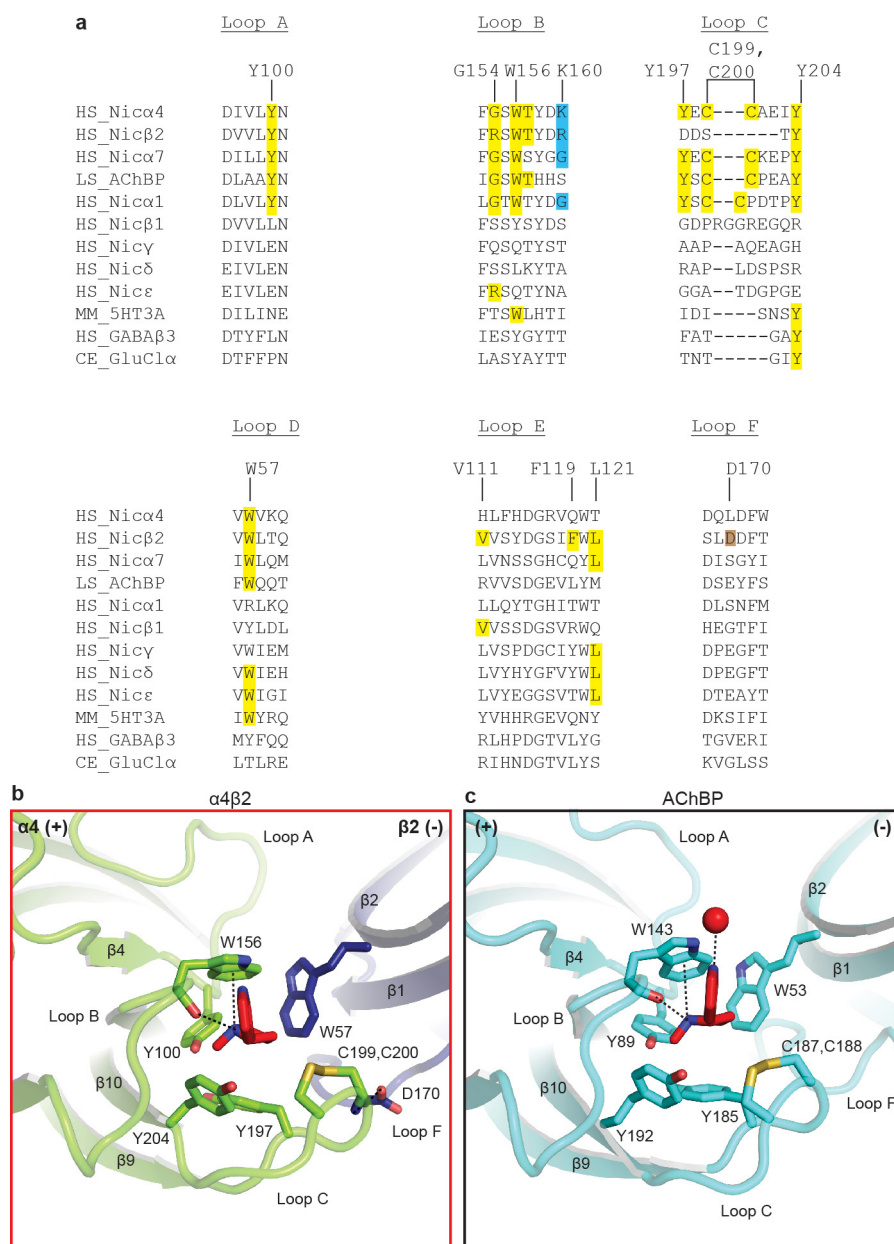
Extended Data Figure 4 | Structural superimpositions. **a**, C α atom r.m.s.d. from pairwise superimpositions of all $\alpha 4$ and $\beta 2$ chains. **b**, Backbone comparison of the $\alpha 4$ (green) and $\beta 2$ (blue) subunits. **c**, Superimpositions of subunits of representative pentameric ligand-gated ion channel structures (magenta) on the chain A $\alpha 4$ subunit (green). PDB accessions and C α r.m.s.d. values are listed. Asterisk indicates bulging caused by inserted leucine residue found in the M2–M3 loops of $\alpha 4$ and $\beta 2$ subunits relative to other receptors shown here (this loop was unmodelled in the 5-HT $_3$ R structure, however, that protein has the same loop length as $\alpha 4$ and $\beta 2$). The most similar subunit structure overall to $\alpha 4$ is GLIC,

which has been thought to represent an open state; however, studies on its desensitization properties^{58–60} and comparison to the $\alpha 4\beta 2$ receptor structure here and in Extended Data Fig. 8 suggest it may rather represent a desensitized conformation. Conversely, the *Torpedo* nicotinic receptor structure, while clearly adopting the same overall fold, aligns less well structurally with $\alpha 4$ than does GLIC. This difference may relate to the *Torpedo* receptor being in a closed-resting state; notable differences in the backbone conformation of the *Torpedo* M2–M3 and Cys-loops (inset) compared to all other structures are less straightforward to interpret.



Extended Data Figure 5 | Detailed interface interactions. **a–c**, Views parallel and perpendicular to the plasma membrane, colouring potential van der Waals (grey), hydrogen bonds (orange) and electrostatic (pink) interactions in the subunits interface. Parallel views are from periphery of receptor. **d–f**, Close-up of the red boxes on the apical receptor surface. **g–i**, Close-up of the black boxes in the view parallel to the plasma membrane. **j–l**, Close-up of the yellow boxes in the view parallel to the plasma membrane. Panels **j–l** highlight the N-capping of the M1 helix by a serine in the M2–M3 loop, an interaction seen in GlyR-closed, but absent in GlyR-open and GABA_AR^{17,24}. For simplicity, only the residues likely to

be involved in forming hydrogen bonds and electrostatic interactions are shown. These potential interactions are shown as dashed lines (2.4–3.9 Å). The subunit interfaces are predominantly stabilized through van der Waals interactions, with interspersed hot spots of hydrogen bonding and electrostatic interactions of known functional importance. The N-terminal helix of the receptor is important in pentameric assembly and mutations in this region of other pentameric receptors results in disease¹⁷. Loop C is essential for orthosteric ligand binding, the M2–M3 loop is critical for allosteric signal transduction⁷, and residues at the apex of M1 and at the intracellular base of the pore are known to affect desensitization^{25,61}.

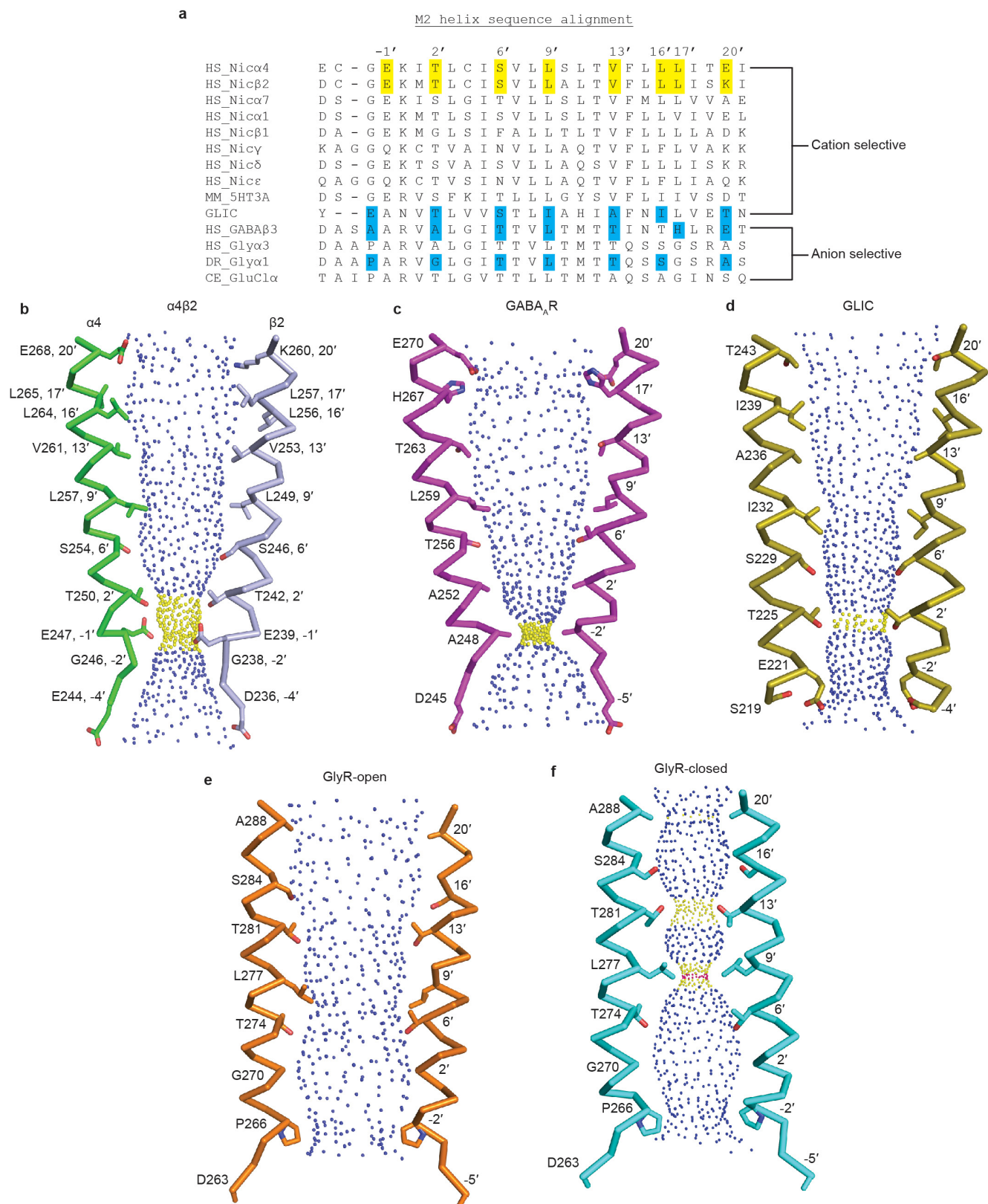


Extended Data Figure 6 | Determinants of nicotine binding.

a, Sequence alignment of loops implicated in nicotine binding. The human nicotinic α1 (NCBI GI accession number 87567783), β1 (41327726), γ (61743914), δ (4557461) and ε (4557463) subunits were added to the sequence alignment. Residues making contact with nicotine or stabilizing the binding pocket indirectly are highlighted in yellow and brown, respectively. Determinants indirectly affecting the

receptor–nicotine cation–π interaction are highlighted in blue.

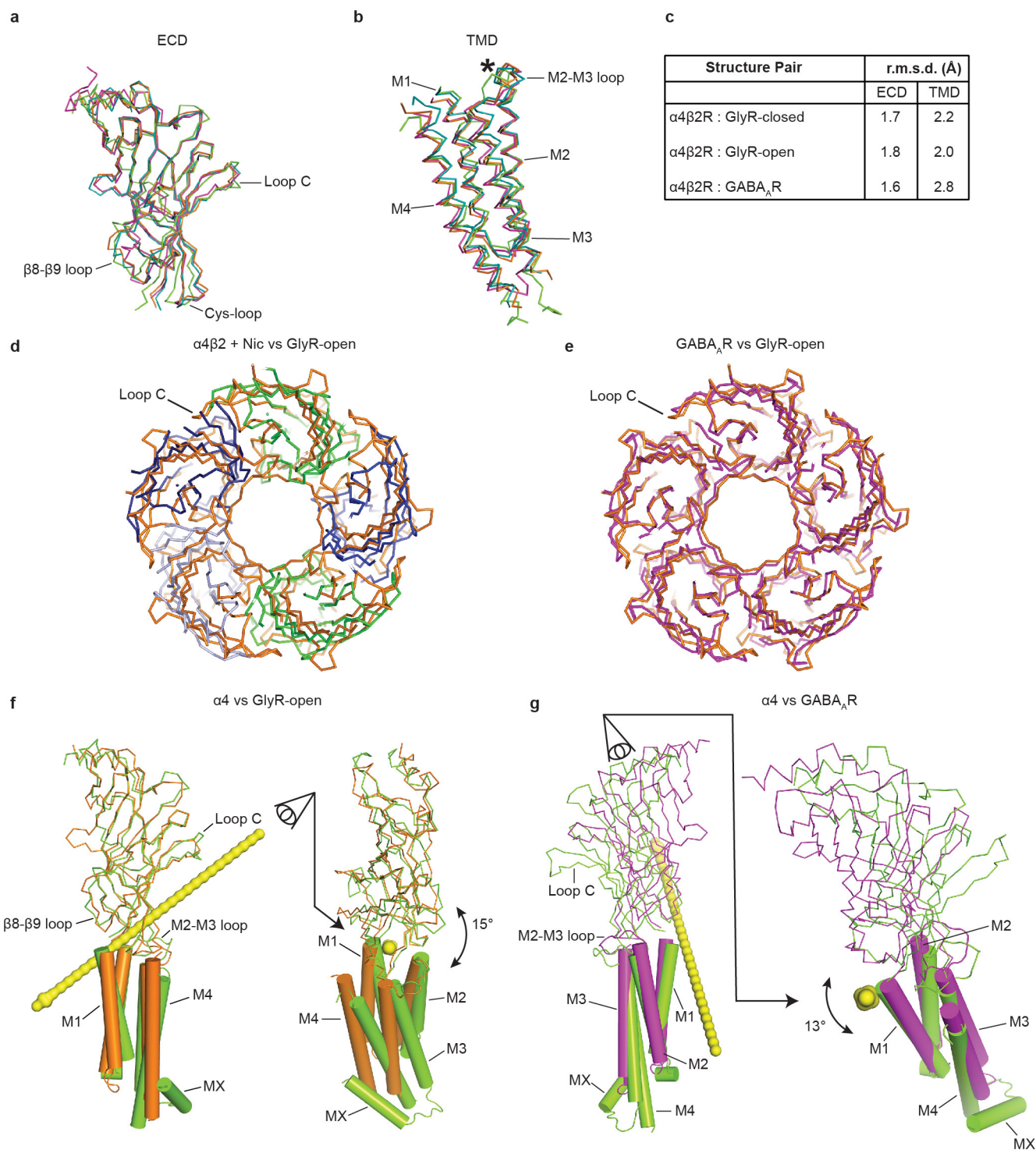
b, Close-up of the α4β2 nicotinic receptor binding pocket. **c**, Close-up of the corresponding region in AChBP (PDB accession 1UW6)⁵⁵. The water in the AChBP pocket is represented as a red sphere and forms a hydrogen bond between the pyridine nitrogen on nicotine and the protein backbone. Potential hydrogen bonding and cation–π interactions are represented as dashed lines (2.7–5 Å).



Extended Data Figure 7 | Cys-loop receptor ion channel conformations.

a, Sequence alignment of the M2 α -helices. Residues lining the α 4 β 2 receptor pore are highlighted in yellow and the residues lining the pores of GlyR (closed: PDB accession 3JAD; open: 3JAE)²⁴, GLIC (4QH5)⁶² and GABA_AR (4COF)¹⁷ are highlighted in blue. **b–e**, View of the M2

α -helices from opposing subunits with side chains shown for pore-lining residues. The blue and yellow spheres represent the internal surface of the transmembrane ion channel. Blue spheres are pore diameters >5.6 Å; yellow are >2.8 Å and <5.6 Å; and pink are <2.8 Å.



Extended Data Figure 8 | Comparison of Cys-loop receptor conformational states.

a, View parallel to the plasma membrane of a superposition of the $\alpha 4$ subunit (green) ECD with the GABA_AR (magenta) and GlyR-open (orange) and GlyR-closed (cyan). **b**, View parallel to the plasma membrane of a superposition of the TMDs. Asterisk indicates an inserted leucine in the M2–M3 loop of $\alpha 4\beta 2$, which is conserved in 5-HT₃ receptors. In the high-resolution structure of the 5-HT₃R, the majority of the M2–M3 loop including the leucine of interest is not modelled, precluding comparison of the two structures for this analysis. **c**, Table of

C α r.m.s.d. values between isolated regions of one subunit per structure.

d, **e**, View down the channel axis from the synaptic cleft towards the cytosol of a superposition of the receptors based on alignments of the TMDs. **f**, **g**, Analysis of intra-subunit rotation angles between different conformational states. Rotation axes indicated by yellow bar. In **f**, the ECD of GlyR-open was superposed on the ECD of $\alpha 4$ and relative displacement of the TMD is shown. In **g**, the TMD of GABA_AR was superposed on the TMD of $\alpha 4$ and relative displacement of the ECD is shown.

Extended Data Table 1 | Data collection and refinement statistics

Dataset	Nicotine	5-Iodo-A-85380*
Data collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Resolution (Å) [†]	40.00-3.94 (4.01-3.94)	30.00-6.50 (6.61-6.50)
Wavelength (Å)	0.9791	1.6984
Cell dimensions a, b, c (Å) [‡]	127.1, 132.6, 202.4	128.1, 133.6, 205.6
Number of unique reflections	30759	7259
Completeness (%) [†]	99.5 (97.8)	99.2 (100)
Redundancy [†]	9.1 (7.5)	6.3 (6.5)
I/σ(I) [†]	14.9 (1.1)	19.4 (1.5)
CC1/2 in the last shell	0.547	0.528
Refinement		
Resolution (Å) [†]	25.00-3.94 (4.08-3.94)	
Number of reflections (test set)	26,718 (1,330)	
Completeness (%) [†]	86.8 (33)	
R _{work} /R _{free} (%)	28.5/30.7	
Number of non-H atoms	14,805	
Mean B factors (Å ²)		
Protein	170	
Ligand/carbohydrate	147	
Water/ion	74	
r.m.s.d. values		
Bond lengths (Å)	0.003	
Bond angles (°)	0.745	
Ramachandran analysis		
Favored (%)	93.8	
Outliers (%)	0	
Molprobability score	2.47 (99 th percentile)	

*This data set is of low resolution and was only used to generate anomalous difference maps.

[†]Values in parentheses are for the highest resolution shell.[‡]All angles = 90°.

Extended Data Table 2 | Surface areas buried at subunit interfaces

a	Structure (PDB ID)	Interface area (Å²)	
		(+) subunit	(-) subunit
	$\alpha 4\beta 2$ [α - β interface]	2820	2806
	$\alpha 4\beta 2$ [β - β interface]	2501	2575
	$\alpha 4\beta 2$ [β - α interface]	2544	2561
	nAChR [α - γ interface] (2BG9)	1665	1658
	nAChR [α - δ interface] (2BG9)	1308	1300
	nAChR [β - α interface] (2BG9)	1426	1401
	nAChR [γ - α interface] (2BG9)	1684	1714
	nAChR [δ - β interface] (2BG9)	1858	1842
	5-HT ₃ R (4PIR)	3125	3012
	GABA _A R (4COF)	2560	2621
	GlyR + gly (3JAE)	1708	1760
	GlyR + strychnine (3JAD)	2155	2137
	GlyR + strychnine (5CFB)	2214	2273
	GluCl (3RHW)	2231	2298
	GLIC (4HFI)	2215	2121
	ELIC (2VL0)	2593	2474

b	Structure (PDB ID)	Loop C interface area (Å²)	
		(+) subunit	(-) subunit
	$\alpha 4\beta 2$ [α - β interface]	234	249
	$\alpha 4\beta 2$ [β - β interface]	31	34
	$\alpha 4\beta 2$ [β - α interface]	31	34

a, Buried area at subunit interfaces in the $\alpha 4\beta 2$ receptor and other pentameric receptors. The 5-HT₃R structure contains an extra section of the intracellular domain (Extended Data Fig. 4c), which accounts for its larger subunit interface area. Glycine receptor structures include two from cryo-electron microscopy studies (PDB accessions 3JAE and 3JAD in the open and resting states, respectively)²⁴ and one from X-ray crystallography in the resting state (PDB accession 5CFB)⁵⁶. **b**, Surface areas buried by only loop C. We analysed inter-subunit interactions in the $\alpha 4\beta 2$ receptor to investigate mechanisms underlying heteromeric receptor assembly. The crystal structure of the receptor reveals three classes of subunit interfaces: α - β , β - β and β - α . All three interface types in the receptor are comparable in terms of surface area buried to the most tightly packed Cys-loop receptor structures. Of the three interface classes in the $\alpha 4\beta 2$ receptor, the α - β interface is the most extensive; the majority of this difference is provided by loop C, which is considerably longer in the α subunit and forms extensive contacts with the neighbouring β subunit (Extended Data Fig. 5g–i). Among the pentameric receptors of known structure, the $\alpha 4\beta 2$ nicotinic receptor is closest in sequence and function to the *Torpedo* nicotinic receptor⁵. We compared backbone conformations and inter-subunit interactions between these two structures (Extended Data Fig. 4c). We found that the $\alpha 4\beta 2$ receptor conformation is more similar to other eukaryotic receptors and the bacterial receptor GLIC than to the *Torpedo* receptor. We also observed that subunit interfaces are much more loosely packed in the *Torpedo* receptor structure. Owing to these differences, and to a previously described register inconsistency in its TMD^{17,24,41–43}, we limited our further structural comparisons with the *Torpedo* nicotinic receptor.

ADDENDUM

doi:10.1038/nature19064

Addendum: Non-Joulian magnetostriction

Harsh Deep Chopra & Manfred Wuttig

Nature **521**, 340–343 (2015); doi:10.1038/nature14459

In this Letter, we showed that the volume of the Fe–Ga crystals we investigated is not conserved in the course of magnetostriction measurements; we termed this phenomenon non-Joulian magnetostriction (NJM), in contrast to Joule magnetostriction, which is volume conserving¹. We measured NJM in circular-shaped single-crystal disks by applying an in-plane magnetic field and showed that the disks expand radially. Magnetostriction normal to the disks was not reported because we assumed that a negligible vector component of magnetization normal to the disk at fields at which NJM is realized would yield negligible magnetostriction.

Here we present precision measurements undertaken to experimentally verify this assumption. The results are represented by the red curve in Fig. 1. We measured NJM with the strain-gauge technique described in our Letter², which uses a Wheatstone bridge combined with lock-in null detection featuring a resolution of 0.2 p.p.m. We attached micro-strain gauges (300- μm gauge length) on the cylindrical surfaces of the samples (lower-right inset in Fig. 1). The data shows a very small strain of 1.3 p.p.m., normal to the disk. In this example, the field was directed parallel to the in-plane $[110]$ axis of the Fe–Ga crystalline disk. Its longitudinal and transverse magnetostriction strains are 70 p.p.m. and 62 p.p.m., respectively, whereas strain along the $[100]$ axis equals 89 p.p.m. (Fig. 1). A similarly negligible strain (1 p.p.m.) occurs when the field is directed along the $[100]$ axis in the plane of the disk (not shown). We also noted that the vector component of magnetization in the $[001]$ direction for an in-plane field (along any in-plane direction) is negligible (upper inset of Fig. 1).

We thus maintain our original conclusion that the disk expands and the volume is not conserved (NJM). In an upcoming paper (R. U. Chandrasena, W. Yang, J. A. Boligitz, M. Forst, A. Scholl, E. Arenholz, F. Kronast, H. Ebert, J. Minár, A. X. Gray & H.D.C., manuscript in preparation) we show that the observed NJM originates from the nanometre-scale lamellar structure within the highly periodic cellular domains shown in Fig. 3 of our original Letter. Degradation of the lamellar or cellular structure causes the disappearance of non-Joulian behaviour. The generalized Landau-type magnetic structure in Fig. 3 of our original Letter has an electronic origin (charge density waves)

and a long coherence. Its existence is a prerequisite of the non-Joulian character of the magnetostriction in Fe–Ga.

We acknowledge the contribution of C. Jiang, Y. He, P. Stamenov, M. Coey and H. Xu for drawing the omission of this data to our attention. H.D.C. acknowledges the support of National Science Foundation DMR-Condensed Matter Physics grant number 1541236 and Temple University OVPR's Infrastructure Grant and Temple University Merit Scholars grants. M.W. acknowledges the support of ARO grant W911NF-15-1-0615.

1. Joule, J. P. On the effects of magnetism upon the dimensions of iron and steel bars. *Phil. Mag. J. Sci.* **30**, 76–87, 225–241 (1847).
2. Sullivan, M. Wheatstone bridge technique for magnetostriction measurements. *Rev. Sci. Instrum.* **51**, 382 (1980).

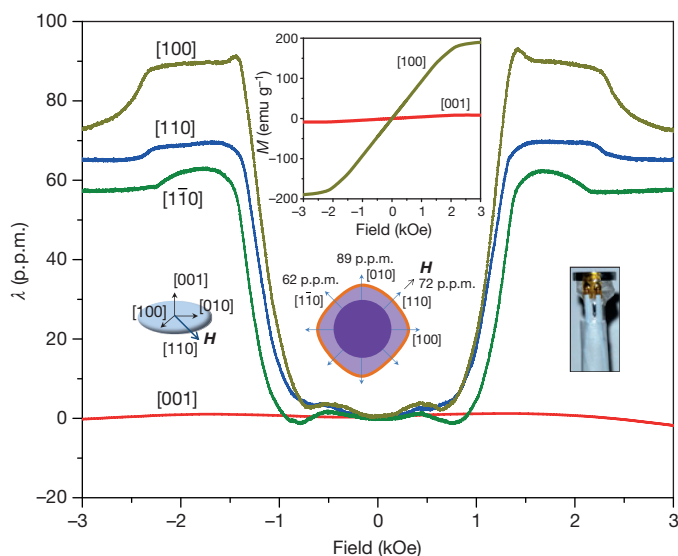


Figure 1 | Volume is not conserved in non-Joulian magnetostriction. Room-temperature magnetostriction λ along various principal directions of a slow-cooled $\text{Fe}_{82.9}\text{Ga}_{17.1}$ single crystal with applied field, H , along a $[110]$ axis. The red curve shows measured magnetostriction along the $[001]$ direction, this direction being normal to the disk, as shown schematically in the lower-left inset. The lower-middle inset shows expansion along all directions in the plane of the disk. The upper inset shows magnetization M along the $[100]$ -type direction along with simultaneously measured orthogonal (vector) data in the $[001]$ direction, the latter being negligible in the field range for which NJM is observed. The lower-right inset shows a photograph (taken by H.D.C.) of the micro-strain gauge setup attached to the cylindrical surface of a sample.

CAREERS

MOBILITY A view beyond your home nation go.nature.com/2dmkdd1

@NATUREJOBS Follow us for the latest news and features go.nature.com/2def08e

NATUREJOBS For the latest career listings and advice www.naturejobs.com

HERO IMAGES/GETTY



Informal chats with people in the know can help researchers to decide whether a career path is likely to be a good fit.

COLUMN

For your information

Sounding out people who are already working in a field that interests you is a great way to gain valuable inside knowledge during your job search, says **Peter Fiske**.

As a scientist, you've learnt to be resourceful and self-reliant — to research your questions and solve problems with little or no help from anyone else. The culture of the scientific enterprise encourages this go-it-alone approach, and the research community often recognizes individual contributions more than it does group efforts. The emphasis on attacking problems single-handedly may serve you well in your research, but it can hold you back when it comes to exploring career options or investigating new pathways in your professional development. For example, many early-career scientists search for jobs without seeking help from a hugely effective career-development tool — informational interviewing.

What exactly is this? It is not a job interview: you are not selling your candidacy to the person you talk with. Rather, you're aiming to learn about that person's job, their remit and their field, and what it's like to work at their organization. The practice is one of the best ways to get inside information about a company, business, non-profit group or other organization where you might wish to work, and about the field or discipline itself. It is a common technique used by professionals outside the academic science enterprise to learn about opportunities, and about employers and industries that are unfamiliar to them. It's not unusual for these people to conduct 20–40 informational interviews in the course of a single job search.

Yet this type of meeting is barely discussed — and seldom encouraged — in academia. Even now, many faculty members are unfamiliar with professional customs beyond their campus, and may be resistant to their students exploring beyond the conventional PhD career pathways. And many PhD programmes still emphasize career paths in research science and academia, for which informational interviewing is not the norm. As a result, young scientists rarely consider, let alone arrange, any such discussion, and this puts them at a disadvantage, especially if they hope to move out of academic research.

Setting up and conducting an informational interview may seem strange and awkward. And it's true that talking to a single person gives ►

► you only a single perspective on a career or an employer. But the real value of an informational interview is learning what that person knows and has observed about their job, the organization that employs them and the field in which they work. And multiply one interview by 10 or 20 and you have that much more information about a variety of employers and industries. Just as importantly, many of the people you meet through such interviews will be willing to help you in your job search, offering advice and introductions. You gain not only insight but often supporters and advocates.

The best candidates for these one-to-one meetings are those who work in the field or for the employer that interests you, and with whom you have a connection, no matter how distant or tangential. Many universities' career-planning and placement centres keep a database of graduates who work in a wide range of disciplines and fields and have volunteered to speak to students and postdocs about their own careers and experience. Aim to find at least one person by this means whose educational background is similar to yours and who works in a field that you'd like to pursue. You can learn extremely useful information about how they made the move from research science into this particular career.

You should also try reaching out to members of your professional network, who will be able to introduce you to someone they know. The personal introductions that your network contacts can make on your behalf will create a crucial first impression with those interview targets.

FIRST STEPS

To set up the informational interview, contact your target candidate through e-mail, introduce yourself and explain who referred you and why you seek an informational interview. You should also briefly describe why you are interested in that person's industry or field and what you hope to learn from the discussion. Don't give up if at first you get no response. A brief and cordial follow-up e-mail is appropriate if you have not received a reply after a week or so. Showing persistence and positivity in seeking this meeting is a good thing: don't be afraid to call the person if you don't get a response from your e-mails.

Once you've scheduled the discussion, you should provide a bit more information about yourself. Don't send a résumé or CV, because it could signal that you're seeking a job rather than information. Instead, send a one-paragraph summary of your background, education, key accomplishments and professional interests. Including the URL to your LinkedIn page will help your interviewee to become more familiar with you before the meeting.

"Many people will be willing to help in your job search. You gain not only insight but often supporters and advocates."

QUESTION TIME

How to get the most from your meeting

Informational interviews are a great way to get direct, candid feedback and advice on potential career paths long before you begin your job search. Here are some questions to ask during the meeting.

● **Why did you make the move from research and what drew you to this career?**

This is a good way to start an interview because it invites the respondent to share their feelings and personal experiences. Although everyone's story is different, you can often uncover common drives and mutual interests in their response.

● **How did you make the transition?**

Your host's answer to this question can help you to understand how someone identified

and researched their options and ultimately made their choice. Listen carefully to what they say about the key sources of information that they used, and the part that networking played in their job search and career transition.

● **If you could go back to graduate school and take one class, or develop one skill that would help you in your present career, what would it be?**

This is a great way to solicit advice about the key skills, knowledge and experience that would make you more competitive for a job in this career field. You may be able to act on the advice immediately, perhaps by seeking out a class or short course on the subject that your interviewee recommends. **P.F.**

Keep things simple for your interviewee: ask to schedule the interview at their workplace at a time of their choice. Not only does this minimize disruption for them, but it also gives you the chance to see that work environment. Sometimes, the person you meet will be happy to show you around and introduce you to others in their organization. But even if that's not the case, you can learn things about their workplace (such as whether people work collaboratively or alone, how they dress and the general ambience) that can help you to decide whether the environment would be a good fit for you.

You should arrive on time for the meeting with a list of prepared questions and topics (see 'Question time'). And in this setting, unlike at a first interview for a specific job, you can freely ask about salary ranges, typical benefits, time off and other such delicate issues. Plan to meet for no longer than 30 minutes. Sometimes, however, these meetings can go so well that neither you nor the interviewee is ready to stop after half an hour. In this case, be courteous and check with your host that it's OK to continue.

It is professional protocol to e-mail your interviewee within 24 hours of the meeting, thanking them for their time and for the insights they shared. If specific follow-up items came out of the interview, such as sending a copy of your résumé or an article you referred to during your discussion, be sure to attend to those quickly. Contact the person again by e-mail 10–12 weeks later. Thank them once more for their help and update them on your career-exploration progress. I know several PhDs who took this tack with every person they had an informational interview with. In at least one case, the 3-month follow-up so impressed the interviewee that it sparked another discussion — which led to a job offer.

If you fear that you may have nothing to

offer the other person, don't use that as justification to avoid setting up a meeting. The person you contact will have sound reasons for wanting to meet you. Many professionals agree to meet because they want to do a favour for (or return one to) the person who introduced you. At times, your target may know that their employer will be hiring soon, and they may want to meet a potential candidate who has already expressed interest in their field or workplace. And sometimes people are motivated simply by kindness or curiosity.

Get comfortable with and embrace this practice: it has value far beyond the job search. Seeking out people who work in fields or organizations of interest and launching conversations with them is a key practice of successful professionals. Sometimes crucial insights and opportunities can emerge from conversations with people who have only the most superficial connection to your current career path. "Chance favours the prepared mind," as Louis Pasteur said. And indeed, the best career opportunities often favour those who invest some time seeking out others and learning from them. ■

Peter Fiske is chief executive of PAX Water Technologies in Richmond, California, and author of *Put Your Science to Work* (American Geophysical Union, 2001).

CORRECTION

The Careers feature 'Going for broke' (*Nature* **534**, 579–581; 2016) conflated the ideas of an emergency account and an emergency fund. The emergency fund would include an emergency account, as well as other subaccounts for unexpected expenses.

THE MOST IMPORTANT THING

You must remember this ...

BY MARISSA LINGEN

What's the most important thing that happened in 2048?

A1: '48? That's when they found the cure for the grouse flu. Don't know what the poultry population would have done without that. It might have spread to the wild birds like the turkey flu did, and without another replacement population ... I don't like to think about hypotheticals. Bad, though. Dodged that bullet.

A2: The *Star Wars Droid Pilots* series started! Man, I gave a decade of my life to that fandom. The first convention wasn't until '49, though, so in '48 I was still going to *Ghostbusters* cons.

A3: Lorelei was born. She was so waxy at first, not red like a baby is supposed to be. She didn't cry for so long — minutes. And then she breathed a little and then this thin little scream, and I knew it would be all right. I can't think of anything that could possibly be more important than Lorelei.

A4: Oh, I know, you could use one of those services that gives you Top Headlines of 2048! You could look it up on your device right now! Or, wait, I'll ask my social hub, I bet they could tell you! I've got 40 answers collated, just a sec, they group into three main topics.

A5: Those single-celled organisms on Europa, that was '48. We worked another ten years on those, through the rough times

without more data. That was enough data to keep us going. Of course, we barely ate, but who cared? Europa! Nobody could give another answer, that's what '48 means.

A6: That was the earthquake in Argentina, wasn't it? I didn't know anybody down there, but it meant the beginning of the Pan-Southern Unification. That's pretty important, I guess. I mean, not for me personally, I don't know much about that sort of thing, for me it was probably that I took up macramé and got my cousin to leave that bastard Pat Schmidt, you don't even want to know. But I try to keep perspective. The rest of the world is out there.

A7: The Lutheran Church-Jefferson Synod broke off from the Lutheran Church-Missouri Synod. It was the most crucial church polity question of our time, I can't believe you're even asking — oh. You're not really asking, you just want to know if I remember what year it was. Of course I remember, it was '48.

A8: Plantain chicken enchiladas at Perez's. Enough said, right? Everyone tried to replicate that recipe, everyone. Even me. With a lager, nothing like it on a summer night, not so bad in the winter either. I think I ate nothing else for weeks that June. That all-spice blend was probably coming out my pores. That's what my wife would tell you, '48, the year he stank of allspice and wouldn't stop trying to figure out which pepper blend.

A9: When my father died in '48, we couldn't keep Mum in the house any more. So we

spent all year moving her into an apartment. And then, well, you can guess how that turned out within a few years, all that work for nothing. God, what a decade, you might as well ask, what's your favourite time you got an ear infection? Most important thing that happened in '48? Christ, take your pick.

A10: My sister Janice graduated. She was the last of us who did. We couldn't afford anyone else's tuition for a while, it was before the reforms went through — you know all that — but we had Janice, my folks and aunts and uncles had all put in so she could get a pharmacy degree, and as you'd expect, *that* was handy. Janice hated it, but nobody much cared what Janice thought by then. Oh, it was a big party. We cooked for days for her graduation. I was jealous as anything. I miss her now.

A11: That was the year of the dingo resettlement, wasn't it? I think it was. I think that was the year Australia got too hot and they had to start moving animals. And they started with the dingos because they thought people would take to them well because they were like wee doggies? Heh, God, humans are dumb sometimes, I wouldn't take us on a bet. It wasn't the kangaroos until '50, was it? Maybe the kangaroos was '48. I'll look it up, shall I? Oh, have you got it. All right. Thanks.

A12: You want me to say President Banks, don't you? Because of the nanobombs? That's the answer you're looking for. I bet all the city people say President Banks got elected, that's the most important thing that happened in '48. Look, I got a new spray for the wheat rust. It held it off another three years. And Rob got a new combine, that was '48, that meant that when the nanobombs came, we were fine, we could hold out until everything got put back together again. You're not going to get the answer you want, all right? It's not all the way you think it is. Everybody else can say President Banks, I don't care, I'll be the only one who doesn't. Sometimes it's the combine. ■

Marissa Lingen has published more than 100 short stories in venues such as *Analog*, *Lightspeed* and *Tor.com*.



ILLUSTRATION BY JACEY